

**МИНИСТЕРСТВО ВНУТРЕННИХ ДЕЛ РОССИЙСКОЙ ФЕДЕРАЦИИ**  
**ФЕДЕРАЛЬНОЕ КАЗЕННОЕ УЧРЕЖДЕНИЕ**  
**«ГЛАВНЫЙ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИЙ ЦЕНТР»**



**Обзор**  
**отдельных вопросов**  
**в области больших данных**  
**и искусственного интеллекта**  
**IV ВЫПУСК**  
**том 2**

**Москва 2020**

Под общей редакцией В.С. Овчинского,  
доктора юридических наук,  
советника Министра внутренних дел  
Российской Федерации

Обзор отдельных вопросов в области  
больших данных и искусственного интеллекта.  
IV выпуск, 2 том.  
М.: ФКУ «ГИАЦ МВД России», 2020. – 315 с.

В обзоре представлены материалы, касающиеся вопросов использования больших данных и развития искусственного интеллекта в мире. Отдельно отражены актуальные вопросы использования ИИ в правоохранительной системе.

## Оглавление

<b>Правосудие по алгоритму – роль искусственного интеллекта в полицейской деятельности и системах уголовного правосудия .....</b>	<b>6</b>
1. Ключевые понятия .....	13
2. Реальные приложения алгоритмов и искусственного интеллекта в системах уголовного правосудия .....	20
3. Надзор и регулирование .....	28
Выводы и рекомендации .....	33
<b>Интерфейс мозг-компьютер: новые права или новые угрозы основным свободам? .....</b>	<b>37</b>
Краткое содержание .....	37
Технологии .....	43
Приложения .....	47
Проблемы .....	54
5 ответов .....	56
Выводы и рекомендации .....	65
<b>Предотвращение дискриминации, вызванной использованием искусственного интеллекта .....</b>	<b>65</b>
Резюме .....	66
Частный сектор .....	74
Государственный сектор .....	77
Частный случай информационных потоков .....	80
Дизайн и назначение .....	84
Разнообразие .....	86
Синдром «чёрного ящика»: прозрачность, объяснимость и подотчётность .....	87
Выводы .....	91
<b>Необходимость демократического управления ИИ .....</b>	<b>93</b>
Введение .....	100
Порядок действий .....	100
Обоснование отчета .....	101

Доступ к информации – дезинформация, «эхо-камеры» и эрозия критического мышления .....	105
Массовая слежка и усиление авторитаризма .....	109
ИИ и принятие политических решений .....	109
<b>Правовые аспекты автономных транспортных средств .....</b>	<b>118</b>
Введение .....	119
Ключ концепции-автоматизация транспортных средств .....	124
Уголовная ответственность .....	129
Гражданская ответственность за ущерб .....	131
Ответственность производителя .....	132
Страховые проблемы .....	134
Текущее регулирование в Европе .....	135
Выводы и рекомендации .....	138
<b>Китайские взгляды на аналитику больших данных .....</b>	<b>140</b>
Резюме .....	142
Глава первая .....	146
Глава вторая .....	149
Глава третья .....	154
Глава четвертая .....	166
Глава пятая .....	182
Заключительные мысли и будущее .....	182
<b>Искусственный интеллект в области юриспруденции .....</b>	<b>185</b>
<b>Статья 1 .....</b>	<b>185</b>
Введение .....	185
1. LegalTech в России .....	186
2. Особенности предметной области .....	188
3. Анализ существующих подходов и инструментов .....	192
4. Выводы по итогам тестирования .....	218
<b>Статья 2 .....</b>	<b>221</b>
1. Legal AI: миф или реальность .....	222
2. Новый подход к созданию продукта .....	225

3. Графовые системы и онтологии .....	239
4. Заключительные положения .....	264
<b>Статья 3 .....</b>	<b>267</b>
Введение.....	267
1. Логическая интерпретация данных .....	268
2. Графы знаний и онтологии.....	277
3. Возможность доработки под особенности отечественного права. ....	302
4. Логика и удобство использования.....	302
5. Ширина охвата правоотношений. ....	302
6. Автогенерация онтологий .....	312
7. Заключительные положения .....	314

## **Правосудие по алгоритму – роль искусственного интеллекта в полицейской деятельности и системах уголовного правосудия**

Отчёт | Док. 15156 | 01 Октябрь 2020

Комитет по правовым вопросам и правам человека

Докладчик: г-н Борис ЧИЛЕВИЧС, Латвия, СОС

Происхождение Обращение в комитет: Док. 14628, номер 4407 от 12 октября 2018. 2020 г.

### **Резюме**

Всё более мощные и влиятельные приложения искусственного интеллекта (ИИ) теперь можно найти во многих сферах человеческой деятельности. Система уголовного правосудия представляет собой одну из ключевых сфер ответственности государства. Это даёт властям значительные полномочия по вмешательству и принуждению. Таким образом, привнесение нечеловеческих элементов в процесс принятия решений в системе уголовного правосудия может создать особые риски.

Если ИИ будет внедряться с согласия общественности, то эффективное и соразмерное регулирование является необходимым условием. Будь то добровольное саморегулирование или обязательное правовое регулирование, это положение должно основываться на общепринятых и применимых основных этических принципах.

Есть много причин, по которым использование ИИ в системах охраны правопорядка и уголовного правосудия может не соответствовать основным этическим принципам. Хотя использование ИИ в системах охраны правопорядка и уголовного правосудия может иметь значительные преимущества, если оно будет надлежащим образом регулироваться, оно рискует оказать особенно серьёзное воздействие на права человека, если это не так.

Поэтому Комитет по правовым вопросам и правам человека предлагает государствам – членам принять ряд мер для обеспечения эффективного регулирования применения ИИ в системах охраны правопорядка и уголовного правосудия. В нём также предлагается, чтобы Комитет министров принял во внимание особенно серьёзное потенциальное воздействие на права человека использования искусственного интеллекта в полицейской деятельности и системах

уголовного правосудия при оценке необходимости и осуществимости европейской правовой базы для искусственного интеллекта.

1. Приложения искусственного интеллекта (ИИ) теперь можно найти во многих сферах человеческой деятельности, от фармацевтических исследований до социальных сетей, от сельского хозяйства до покупок в Интернете, от медицинской диагностики до финансов и от музыкальной композиции до уголовного правосудия. Они становятся все более могущественными и влиятельными, а общественность часто не знает, когда, где и как они используются.

2. Система уголовного правосудия представляет собой одну из ключевых сфер ответственности государства, обеспечивая общественный порядок и предотвращая нарушения различных основных прав путём выявления, расследования, судебного преследования и наказания уголовных преступлений. Он даёт властям значительные полномочия по вмешательству и принуждению, включая наблюдение, арест, обыск и изъятие, задержание и применение физической силы и даже смерти. Не случайно международное право в области прав человека требует судебного надзора за всеми этими полномочиями: эффективного, независимого, беспристрастного контроля за осуществлением властями полномочий в области уголовного права, которые могут серьёзно нарушить основные права человека. Таким образом, привнесение нечеловеческих элементов в процесс принятия решений в системе уголовного правосудия может создать особые риски.

3. Если общественность хочет принять использование ИИ и пользоваться потенциальными преимуществами, которые может принести ИИ, она должна быть уверена в том, что любые риски управляются должным образом. Если ИИ будет внедряться с осознанного согласия общественности, как можно было бы ожидать в условиях демократии, то эффективное и соразмерное регулирование является необходимым условием.

4. Регулирование ИИ, будь то добровольное саморегулирование или обязательное правовое регулирование, должно основываться на общепринятых и применимых основных этических принципах. Парламентская ассамблея считает, что эти принципы можно сгруппировать под следующими общими заголовками:

- 4.1. Прозрачность, в том числе доступность и объяснимость;
- 4.2. Справедливость и беспристрастность, включая недискриминацию;
- 4.3. Человеческая ответственность за решения, включая ответственность и доступность средств правовой защиты;
- 4.4. Безопасность и защита;
- 4.5. Конфиденциальность и защита данных.

5. Ассамблея приветствует Рекомендацию Комитета министров Rec/CM (2020) 1 о влиянии алгоритмических систем на права человека, а также сопровождающие её рекомендации по устранению воздействия алгоритмических систем на права человека, а также рекомендацию Комиссара Совета Европы по вопросам Права человека на тему «Распаковка искусственного интеллекта: 10 шагов для защиты прав человека». Он одобряет общие предложения, содержащиеся в этих текстах, для применения также в области полицейской деятельности и систем уголовного правосудия.

6. Ассамблея отмечает, что во всем мире разработано большое количество приложений искусственного интеллекта для использования полицией и системами уголовного правосудия. Некоторые из них уже использовались или их внедрение рассматривается в государствах – членах Совета Европы. Они включают в себя распознавание лиц, предикативную работу полиции, идентификацию потенциальных жертв преступлений, оценку рисков при принятии решений о предварительном заключении, вынесении приговора и условно-досрочном освобождении, а также выявление «нераскрытых дел», которые теперь могут быть решены с использованием современных технологий судебной экспертизы.

7. Ассамблея считает, что есть много причин, по которым использование ИИ в системах охраны правопорядка и уголовного правосудия может не соответствовать вышеупомянутым основным этическим принципам. Особое беспокойство вызывают следующие системы:

7.1. Системы ИИ могут предоставляться частными компаниями, которые могут полагаться на свои права интеллектуальной собственности, чтобы отказать в доступе к исходному коду. Компания может даже получить право собственности на данные, обрабатываемые системой, в ущерб государственному органу, который пользуется её услугами. Пользователи и субъекты системы могут не получать информацию или объяснения, необходимые для базового понимания её работы. Некоторые процессы, вовлечённые в работу системы ИИ, могут быть не полностью понятны человеку. Такие соображения поднимают вопросы прозрачности (и, как следствие, ответственности/подотчётности).

7.2. Системы искусственного интеллекта обучаются на массивных наборах данных, которые могут быть испорчены историческими предубеждениями, в том числе посредством косвенной корреляции между определёнными прогностическими переменными и дискриминационной практикой (например, почтовый индекс является косвенным идентификатором для этнического сообщества, исторически подвергавшегося дискриминации). Это вызывает особую озабоченность в отношении полицейской деятельности и уголовного правосудия, поскольку в данном контексте широко распространена дискриминация по различным признакам, а также из-за важности решений, которые могут быть приняты. Очевидная механическая объективность ИИ может скрыть эту предвзятость («промывку

технологий»), усилить и даже увековечить её. Некоторые методы искусственного интеллекта не могут быть легко оспорены субъектами их применения. Такие соображения поднимают вопросы справедливости и беспристрастности.

7.3. Ограниченность ресурсов, нехватка времени, непонимание, уважение к рекомендациям системы ИИ, или нежелание отклоняться от них могут привести к тому, что полицейские и судьи станут чрезмерно полагаться на такие системы, фактически отказываясь от своих профессиональных обязанностей. Подобные соображения поднимают вопросы ответственности за принятие решений.

7.4. Эти соображения также влияют друг на друга. Отсутствие прозрачности в приложении AI снижает способность пользователей принимать полностью обоснованные решения. Отсутствие прозрачности и неопределённая человеческая ответственность подрывают способность механизмов надзора и исправления положения обеспечивать справедливость и беспристрастность.

7.5. Применение систем ИИ в отдельных, но связанных контекстах, особенно разными агентствами, последовательно полагающимися на работу друг друга, может иметь неожиданные, даже непредвиденные кумулятивные последствия.

7.6. Добавление элементов на базе искусственного интеллекта к существующим технологиям также может иметь последствия непредвиденной или непреднамеренной гравитации.

8. Ассамблея приходит к выводу, что, хотя использование ИИ в системах охраны правопорядка и уголовного правосудия может иметь значительные преимущества, если оно должным образом регулируется, оно рискует оказать особенно серьёзное воздействие на права человека, если это не так.

9. Поэтому Ассамблея призывает государства – члены в контексте систем охраны правопорядка и уголовного правосудия:

9.1. Принять национальную правовую базу для регулирования использования ИИ на основе основных этических принципов, упомянутых выше;

9.2. Вести реестр всех приложений ИИ, используемых в государственном секторе, и обращаться к нему при рассмотрении новых приложений, чтобы определять и оценивать возможные кумулятивные воздействия;

9.3. Гарантировать, что ИИ служит общим целям политики, и что цели политики не ограничиваются областями, в которых может применяться ИИ;

9.4. Обеспечить наличие достаточной правовой основы для каждого приложения AI и для обработки соответствующих данных;

9.5. Обеспечить, чтобы все государственные органы, внедряющие приложения ИИ, обладали внутренними знаниями, способными оценивать и консультировать по внедрению, эксплуатации и влиянию таких систем;

9.6. Конструктивно консультироваться с общественностью, включая организации гражданского общества и представителей сообщества перед тем, как внедрять приложения ИИ;

9.7. Обеспечивать, чтобы каждое новое применение ИИ было обоснованным, его цель была указана и эффективность подтверждена до ввода в действие, с учётом конкретного операционного контекста;

9.8. Проводить первоначальную и периодическую прозрачную оценку воздействия приложений ИИ на права человека, чтобы, среди прочего, оценивать проблемы конфиденциальности и защиты данных, риски предвзятости/дискриминации и последствия для человека решений, основанных на работе ИИ, с особым вниманием положению меньшинств и уязвимых и обездоленных групп;

9.9. Обеспечивать, чтобы основные процессы принятия решений в приложениях ИИ были понятны их пользователям и тем, на кого влияет их работа;

9.10. Внедрять только приложения ИИ, которые можно изучать и тестировать прямо на месте работы;

9.11. Внимательно рассмотреть возможные последствия добавления элементов на основе ИИ к существующим технологиям;

9.12. Установить эффективные, независимые механизмы этического надзора за внедрением и эксплуатацией систем искусственного интеллекта;

9.13. Гарантировать, что внедрение, функционирование и использование приложений ИИ может быть предметом эффективного судебного надзора.

## **В. Проект рекомендации**

1. Парламентская ассамблея ссылается на свою резолюцию (2020), озаглавленную «Правосудие с помощью алгоритмов – роль искусственного интеллекта в полицейской деятельности и системах уголовного правосудия», которая была принята во время работы в рамках Совета Европы Специальным комитетом по искусственному интеллекту (САНАИ).

2. Ассамблея напоминает, что на все государства — члены Совета Европы распространяются одни и те же базовые стандарты прав человека и верховенства закона, в частности, те, которые установлены Европейской конвенцией о правах

человека (ETS No. 005), в толковании прецедентного права. Европейского суда по правам человека. Он считает, что нормативные лоскутные изделия - различные стандарты в разных странах - могут привести к «этическим покупкам», что приведёт к перемещению разработки и использования ИИ в регионы с более низкими этическими стандартами.

3. Поэтому Ассамблея призывает Комитет министров принять во внимание особо серьёзное потенциальное воздействие на права человека использования искусственного интеллекта в полицейской деятельности и системах уголовного правосудия при оценке необходимости и осуществимости европейской правовой базы для искусственного интеллекта.

## **С. Пояснительный меморандум г-на Цилевича, докладчика**

### **Введение**

1. Предложение о рекомендации, лежащее в основе этого отчёта, которое я представил 26 сентября 2018 года, было передано в Комитет по юридическим вопросам и правам человека 12 октября 2018 года, после чего комитет назначил меня докладчиком 21 января 2019 года. слушания с экспертами на встрече в Берлине 14-15 ноября 2019 г. с участием доктора Майкла Вила, преподавателя цифровых прав и регулирования, Университетский колледж Лондона, и г-жи Марион Освальд, старшего научного сотрудника вице-канцлера по праву, Университет Нортумбрии, Объединённое Королевство. Визит в полицию Уэст-Мидлендса, Соединённое Королевство, был отменен из-за пандемии COVID-19, но был заменён видеоконференциями с Томом Макнилом из Комитета по этике полиции Уэст-Мидлендса и комиссара по преступности и старшим суперинтендантом Крисом Тоддом и Детектив - старший инспектор Ник Дейл из полиции Уэст-Мидлендса. Я хотел бы поблагодарить всех, кого это касается, за их вклад в этот отчёт.

2. Искусственный интеллект (ИИ) больше не является предметом научной фантастики, даже если он ещё не оправдал всех предсказаний научной фантастики. У нас нет разумных машин, способных сопоставить или даже превзойти людей в нескольких областях (так называемый «общий» или «сильный» ИИ). Однако у нас есть системы, способные выполнять определённые задачи, такие как распознавание шаблонов или категорий или прогнозирование поведения, с определённой степенью того, что можно было бы назвать «автономией» («узкий» или «слабый» ИИ). Эти системы можно найти во многих сферах человеческой деятельности, от фармацевтических исследований до социальных сетей, от сельского хозяйства до покупок в интернете, от медицинской диагностики до финансов и от музыкальной

композиции до уголовного правосудия. Они становятся все более могущественными и влиятельными, и общественность часто не знает, когда, где и как они используются. Действительно, чем «лучше» они становятся, тем менее очевидными они могут быть: исследовательская компания Open AI недавно объявила, что воздержится от выпуска новой системы искусственного интеллекта для обработки естественного (то есть человеческого) языка, поскольку она способна создавать текст, который был неотличимы от текста, созданного человеком, и могут быть слишком легко использованы.

3. Система уголовного правосудия представляет собой одну из ключевых сфер ответственности государства, обеспечивая общественный порядок и предотвращая нарушения различных основных прав путём выявления, расследования, судебного преследования и наказания уголовных преступлений. Это даёт властям значительные полномочия по вмешательству и принуждению, включая наблюдение, арест, обыск и изъятие, задержание и применение физической и даже смертельной силы. Не случайно международное право в области прав человека требует судебного надзора за всеми этими полномочиями: эффективного, независимого, беспристрастного контроля за осуществлением властями полномочий в области уголовного права, которые могут серьёзно нарушить основные права человека. Таким образом, привнесение нечеловеческих элементов в процесс принятия решений в системе уголовного правосудия может создать особые риски.

4. Парламентская ассамблея уже затрагивала некоторые вопросы, относящиеся к настоящему отчёту, в своей Рекомендации 2102 (2017) «Технологическая конвергенция, искусственный интеллект и права человека». Он отметил, что «законодателям все труднее адаптироваться к темпам развития науки и технологий и разрабатывать необходимые правила и стандарты». Ассамблея пришла к выводу, что «защита человеческого достоинства в 21 веке подразумевает разработку новых форм управления, новых форм открытых, информированных и состязательных общественных дебатов, новых законодательных механизмов и, прежде всего, налаживание международного сотрудничества, позволяющего решать эти новые проблемы наиболее эффективно».

5. Хотя это не первый случай, когда комитет рассматривает ИИ, я сначала рассмотрю некоторые основные и общие вопросы, опираясь на работу Комитета по культуре, науке, образованию и СМИ при подготовке Рекомендации 2102, прежде чем рассматривать конкретные случаи использования искусственного интеллекта и алгоритмов в системах уголовного правосудия.

## 1. Ключевые понятия

6. Выражение «искусственный интеллект» было впервые предложено в 1955 году Джоном Маккарти (и другими) из Дартмутского колледжа, штат Нью-Гэмпшир, в предложении об исследовательском проекте, основанном на рабочей гипотезе о том, что «каждый аспект обучения или любая другая характеристика интеллекта в принципе может быть описано настолько точно, что можно создать машину, чтобы моделировать это». К сожалению, до сих пор нет общепризнанного определения ИИ, и есть много разногласий по поводу того, как его следует определять. Примечание. Центральным аспектом этой проблемы является отсутствие общего определения человеческого интеллекта - если предположить, что ИИ должен примерно пониматься как имитирующий все или определённые характеристики человеческого интеллекта, которые по определению не будут подходить для любой будущей формы «суперинтеллектуального» общего/сильного ИИ. Примечание. Общее описание ИИ см. в приложении к настоящему отчёту.

7. Иногда говорят, что отличительной чертой ИИ является его «автономность». Однако к этой концепции следует относиться с большой осторожностью, поскольку она может иметь серьёзные последствия, касающиеся подотчётности и ответственности, вплоть до вопроса о том, следует ли считать ИИ моральным агентом или даже юридическим лицом. Это хорошо описано Европейской группой по этике в науке и новых технологиях. «Термин» автономия «происходит из философии и относится к способности людей принимать законы для себя, формулировать, думать и выбирать нормы, правила и законы, которым они должны следовать. Автономия в этически значимом смысле слова может поэтому только быть отнесены к людям. Поэтому неправильно применять термин «автономия» к простым артефактам, хотя и очень продвинутым сложным адаптивным или даже «интеллектуальным» системам. Поскольку ни один артефакт или система – какими бы продвинутыми и сложными они ни были – сами по себе не могут быть названы «автономными». «в исходном этическом смысле они не могут соответствовать моральному статусу человеческой личности и унаследовать человеческое достоинство. Человеческие существа должны иметь возможность определять, каким ценностям служит технология, что является морально значимым и какие конечные цели и концепции достойны того, чтобы за ними следовали. Это нельзя оставлять машинам, какими бы мощными они ни были».

## **1.1 Возможности и риски при использовании ИИ – некоторые общие соображения**

8. Было сказано, что ИИ «может переопределить работу или улучшить условия труда для людей и снизить потребность в человеческом вкладе, вкладе и вмешательстве во время работы. Он может помочь людям или заменить их интеллектуальными технологиями в сложной, грязной, унылой или опасной работе и даже за её пределами». Примечание. Действительно, как отмечалось в пункте 2, ИИ уже применяется во многих областях, чаще всего с положительными результатами.

9. Потенциальная сила ИИ также несёт в себе риски. Его скорость, сложность и масштабируемость означают, что он значительно превосходит людей при выполнении определённых задач. Потенциальная непостижимость самогенерируемых алгоритмов означает, что метод и рассуждения, используемые для получения определённого результата, могут быть неизвестны даже разработчику ИИ. Было высказано предположение, что «AI/AS будет выполнять задачи, которые намного сложнее и эффективнее, чем предыдущие поколения технологий, особенно с системами, которые взаимодействуют с физическим миром, тем самым повышая потенциальный уровень вреда, который может нанести такая система. Примечание. Этот аргумент можно развить дальше: поскольку «продвинутый искусственный интеллект может представлять собой глубокое изменение в истории жизни на Земле, им следует планировать и управлять им с соразмерной заботой и ресурсами».

## **1.2 Регулирование ИИ – некоторые общие соображения**

10. О том, нужно ли, когда и как регулировать искусственный интеллект, написано очень много. Некоторые комментаторы считают, что регулирование нежелательно, поскольку оно сдерживает инновации или побуждает компании ИИ «делать покупки с соблюдением этических норм»; преждевременно, поскольку технология все ещё развивается; и даже невозможно из-за внутренней природы ИИ. При рассмотрении этих вопросов стоит сначала задуматься над прошлым опытом, в частности, с историей регулирования Интернета.

11. В 1996 году, когда использование Интернета начало распространяться, а сегодняшние гигантские интернет-компании только зарождались или ещё не были созданы, Джон Перри Барлоу представил «Декларацию независимости киберпространства» на Всемирном экономическом форуме в Давосе, Швейцария. «Правительства индустриального мира, вы, уставшие гиганты из плоти и стали, я пришёл из Киберпространства, нового дома Разума. От имени будущего я прошу

прошлое оставить нас в покое. Тебя среди нас не ждут. У вас нет суверенитета там, где мы собираемся. Я заявляю, что глобальное социальное пространство, которое мы строим, естественно, независимо от тирании, которую вы пытаетесь навязать нам. У вас нет морального права управлять нами, и у вас нет никаких методов принуждения, которых у нас есть веские основания опасаться».

12. Хотя это заявление в первую очередь предназначалось для людей, его космополитический, либертарианский подход с тех пор был типичным для многих интернет-компаний. До недавнего времени поведение этих компаний не подвергалось серьёзным возражениям со стороны национальных властей, которые не могли или отказывались признать его последствия. Как сказал профессор Пол Немитц, «эта история неспособности распределять ответственность и брать на себя ответственность в эпоху Интернета, как законодателями, так и техническими корпорациями привела к провалам Интернета в форме распространения массовой слежки и вербовки терроризму, подстрекательству к расовой и религиозной ненависти и насилию, а также множеству других катастроф для демократии, последними из которых стали скандал с Cambridge Analytica и рост популистов, часто наиболее искусственных бенефициаров помощи Facebook, YouTube, Twitter и т.д., совмещая рекламные и сетевые методы целевой рекламы, разработанные для получения прибыли, с политической пропагандой».

13. Помимо прошлого опыта, на котором можно поучиться, вопрос регулирования ИИ также можно отличить по степени его развития. Возвращаясь к профессору Немитцу: «ИИ сейчас – в отличие от Интернета – с самого начала – это не младенческая инновация, выдвинутая в основном академиками и идеалистами, а в значительной степени разработанная и внедрённая под контролем самых могущественных корпораций в области интернет-технологий» Другими словами, мы уже достаточно знаем об ИИ на практике, чтобы регулировать его применение, и мы также достаточно знаем о компаниях, которые его используют, чтобы задаться вопросом, может ли обязательное, обязательное регулирование, в отличие от добровольного саморегулирования, основанного на этике быть необходимым.

### **1.3 Регулирование и общественное доверие**

14. Если общественность должна принять использование ИИ и воспользоваться потенциальными преимуществами, которые он может принести, она должна быть уверена в том, что любые риски управляются должным образом. Два ведущих исследователя в этой области отметили: «Мы знаем, что не существует «формулы» для построения доверия, но мы знаем по опыту, что технологиям, как правило, доверяют, если они приносят пользу, безопасны и хорошо регулируются». Искусственный интеллект должен быть невольно или насильственно навязан

широкой публике – другими словами, если он должен быть введён с согласия общественности – тогда эффективное, пропорциональное регулирование является необходимым, хотя и недостаточным условием.

#### **1.4 Нормативная согласованность и гармонизация**

15. Если необходимо ввести регулирование, оно должно быть последовательным и гармонизированным в максимально широком масштабе. Европейская группа по этике в науке и новых технологиях, например, обратила внимание на «риски, присущие нескоординированным, несбалансированным подходам к регулированию ИИ и «автономных» технологий. Регулятивные лоскутные изделия могут привести к «покупкам с этической точки зрения», что приведёт к переносу разработки и использования ИИ в регионы с более низкими этическими стандартами. Если позволить доминировать в дебатах определённым регионам, дисциплинам, демографии или отраслевым субъектам, рискует исключить более широкий набор общественных интересов и перспектив».

#### **1.6 Этическое управление**

16. Даже те, кто считает, что введение правового регулирования ИИ может быть преждевременным, согласятся с тем, что как минимум необходимо этическое управление. Это было определено как «набор процессов, процедур, культур и ценностей, направленных на обеспечение высочайших стандартов поведения. Таким образом, этическое управление выходит за рамки просто хорошего (а именно эффективного) управления, поскольку оно прививает этическое поведение как отдельным дизайнерам, так и организациям, в которых они работают».

#### **1.7 Этические принципы**

17. Идентификация этических принципов необходима, независимо от того, должны ли они лечь в основу этических «кодексов поведения» или также использоваться для установления правовых норм. Исследования показали, что существует широкое согласие по основному содержанию этических принципов, которые должны применяться к системам искусственного интеллекта, в частности по следующим:

*Прозрачность.* Принцип прозрачности можно широко интерпретировать, включая доступность и объяснимость системы ИИ, другими словами, возможности для человека понять, как система работает и как она даёт свои результаты.

*Справедливость и честность.* Этот принцип включает недискриминацию, беспристрастность, последовательность и уважение разнообразия и плюрализма. Это также подразумевает возможность для субъекта работы системы ИИ оспаривать результаты с возможностью исправления и возмещения.

*Обязанность.* Этот принцип включает в себя требование о том, что человек должен нести ответственность за любое решение, затрагивающее права и свободы человека, с определённой ответственностью и юридической ответственностью за эти решения. Таким образом, этот принцип тесно связан с принципом справедливости и беспристрастности.

*Безопасность и охрана.* Это означает, что системы ИИ должны быть надёжными, защищёнными от внешнего вмешательства и безопасными от выполнения непреднамеренных действий в соответствии с принципом предосторожности.

*Конфиденциальность.* Хотя уважение прав человека в целом можно рассматривать как неотъемлемую часть принципов справедливости и беспристрастности, а также безопасности и защищённости, право на неприкосновенность частной жизни особенно важно там, где система искусственного интеллекта обрабатывает личные или частные данные. Поэтому системы искусственного интеллекта должны соблюдать обязательные стандарты Общего регламента защиты данных (GDPR) Европейского Союза и Конвенции Совета Европы о защите частных лиц в отношении автоматической обработки личных данных (ETS № 108) и поправок к ней. Протокол (CETS № 223, «модернизированная» Конвенция 108+), если применимо.

18. В декабре 2018 года Европейская комиссия по эффективности правосудия (CEPEJ) Совета Европы приняла Европейскую этическую хартию использования искусственного интеллекта в судебных системах и их окружении, в которой показано, как ключевые элементы из приведённого выше списка применимы в этом контексте. Европейская этическая хартия CEPEJ устанавливает пять принципов: уважение основных прав; недискриминация; качество и безопасность; прозрачность, беспристрастность и справедливость (включая доступность и понятность - аналогично объяснимости); и «под контролем пользователя».

## 1.8 Этические проблемы в связи с использованием ИИ в системах уголовного правосудия

19. Подробный анализ работы HART (системы на основе ИИ, используемой Полицией Дарема – см. Ниже) привёл к ряду наблюдений и выводов более общего характера. Примечание. Влияние ИИ следует рассматривать в его рабочем контексте, который в случае полиции включает процедуры, цели и процессы принятия решений. Действия полиции как органов государственной власти могут быть предметом судебного надзора, но ИИ не разрабатывает алгоритм с целью сделать его доступным для понимания людьми. Полиция Дарема намеренно выбрала вариант алгоритма HART, который предпочитал «осторожные ошибки», компромисс между (большим) высоким риском ложных срабатываний и (меньшим количеством) ложноотрицательных результатов («опасными ошибками»). «Оправдывает ли общая выгода для общества от конкретного оценочного суждения, встроенного в алгоритм, возможные негативные последствия для отдельных лиц, может в значительной степени зависеть от серьёзности этих последствий». В алгоритм HART также поступают данные только из полиции Дарема, а не из других местных агентств или других полицейских сил. Люди, принимающие решения, будут иметь доступ к более разнообразным источникам информации, поэтому за алгоритмом никогда не должно быть последнего слова. При этом не исключалась возможность того, что некоторые офицеры «(сознательно или иным образом) предпочтут отказаться от ответственности за то, что является рискованным решением алгоритма, что приведёт к десклейнгу и «атрофии осуждения». Такая ситуация со временем также будет иметь негативные последствия. Человек, принимающий решения, может быстро адаптироваться к изменяющемуся контексту, в отличие от алгоритма, которому, следовательно, потребуется «тщательное и постоянное изучение используемых предикторов и частое обновление алгоритма более свежими историческими данными».

20. Во время слушаний в комитете доктор Вил затронул вопрос об использовании ИИ полицией в более широком социальном и политическом контексте. Он отметил, что сокращение услуг социальной помощи, например, может вызвать проблемы, требующие последующего вмешательства других ведомств, включая полицию, собственные ресурсы которой также были ограничены. ИИ был представлен как имеющий преимущества в экономии средств, и поэтому политика контроля может отдавать приоритет подходам на основе ИИ. Разработчики ИИ могут стремиться избегать дискриминационных или других нежелательных результатов, но они не имеют в виду более широкий социальный контекст. Таким образом, предлагаемые ими решения определяют параметры проблемы в зависимости от того, что технически возможно. Это равносильно приватизации политики и могло бы усугубить социальное разделение и усилить несправедливость.

Государственный сектор, включая полицию, поэтому нуждался в собственном внутреннем опыте, чтобы оценить фактическую полезность решений на основе ИИ и быть готовым сказать «нет».

21. Один из коллег г-жи Освальд по этическому комитету полиции Уэст-Мидлендса и комиссара по преступности (см. Ниже), Джейми Грейс, отмечает, что в Великобритании «как предсказывал Дэвид Лайон в 2007 году, «состояние безопасности» отталкивает государство от общественного дискурса и государственной политики, а «состояние безопасности», озабоченное общественной защитой выше любого рассмотрения индивидуальной автономии, «во многом зависит от данных наблюдения». г-н Грейс считает, что AI «обострит и разожжёт напряжённость в области прав человека, которая уже существует в системе уголовного правосудия. Эти неотъемлемые проблемы прав человека включают проблемы конфиденциальности, сдерживание свободы выражения мнений, проблемы, связанные с возможностью расовой дискриминации, и права жертв преступлений на достойное обращение».

22. С тех пор, как я представил свой вступительный меморандум в апреле 2019 года, возникла повсеместная и растущая негативная реакция на использование искусственного интеллекта в полицейской деятельности и системах уголовного правосудия. Я рассмотрю некоторые конкретные ситуации ниже, но на данном этапе стоит рассмотреть основные положения заявления, сделанного рядом высокопоставленных американских исследователей и учёных в июле 2019 года:

«Актуарные досудебные оценки рисков страдают серьёзными техническими недостатками, которые подрывают их точность, достоверность и эффективность. Они не позволяют точно измерить риски, которые судьи должны учитывать по закону. При прогнозировании бегства и опасности многие инструменты используют неточные и слишком широкие определения этих рисков. При прогнозировании насилия ни один из доступных сегодня инструментов не может адекватно отличить риск насилия для одного человека от другого. Вводящие в заблуждение ярлыки риска скрывают неопределённость этих высоких прогнозов и могут побудить судей переоценить риск и распространённость насилия до суда. Для создания прогнозов оценки рисков основываются на глубоко ошибочных данных, таких как исторические записи арестов, обвинений, обвинительных приговоров и приговоров. Эти данные не являются ни надёжным, ни нейтральным показателем скрытой преступной деятельности. Десятилетия исследований показали, что афроамериканцы и латиноамериканцы за одно и то же поведение с большей вероятностью будут арестованы, привлечены к ответственности, осуждены и приговорены к более суровым наказаниям, чем их белые собратья. Оценка риска, включающая эти искажённые данные, приведёт к искажённым результатам. Эти проблемы не могут быть решены с помощью технических исправлений. Мы настоятельно рекомендуем обратиться к другим реформам».

## **2. Реальные приложения алгоритмов и искусственного интеллекта в системах уголовного правосудия**

23. В этом разделе я в основном сосредоточусь на трёх примерах использования ИИ в системах уголовного правосудия: PredPol, который предсказывает, где могут произойти преступления, и на этой основе рассчитывает, как лучше всего распределять ресурсы полиции; HART (Инструмент оценки риска вреда), который прогнозирует риск повторного совершения преступления при принятии решения о возбуждении уголовного дела; и COMPAS (Профилирование исправительных правонарушителей для альтернативных санкций), который также (в основном) предсказывает, будет ли человек совершать повторное преступление. Все три являются примерами «чёрных ящиков» – закрытых систем, внутренняя работа которых не является общедоступной.

### **2.1 Прогнозирование преступности и распределение ресурсов полиции – PredPol**

24. PredPol, калифорнийская компания, выросшая из проекта Калифорнийского университета (UCLA) и Департамента полиции Лос-Анджелеса, определяет «прогнозирующую полицейскую работу [как] практику определения времени и места, где наиболее вероятно совершение конкретных преступлений, а затем патрулировать эти районы, чтобы не допустить совершения этих преступлений». PredPol использует исторические данные клиентского полицейского управления за период от двух до пяти лет для обучения алгоритму машинного обучения, который впоследствии обновляется ежедневно. Используются только три точки данных: тип преступления, место и дата / время. Согласно PredPol, «никакая демографическая, этническая или социально-экономическая информация никогда не используется. Это исключает возможность нарушения неприкосновенности частной жизни или гражданских прав, наблюдаемых при использовании других моделей правоохранительной деятельности, основанных на разведке или прогнозировании». Однако последнее утверждение оспаривается.

25. Эта технология стоит недёшево. Полиция Кента в Соединённом Королевстве использовала PredPol с декабря 2012 года по март 2018 года по цене 100 000 фунтов стерлингов в год. Сообщалось, что во время первоначального четырёхмесячного судебного разбирательства использование PredPol привело к снижению уличной преступности на 6%. Однако в конечном итоге полиция Кента заявила, что, хотя «Predpol имел хорошую репутацию в прогнозировании того, где могут быть совершены преступления место, более сложной задачей является

показать, что с помощью этой информации мы смогли снизить уровень преступности». Тем не менее, полиция Кента была достаточно впечатлена потенциалом этой технологии, и теперь намеревается разработать свою собственную систему.

26. Несмотря на утверждение PredPol об отсутствии возможности нарушения прав человека, систему критиковали за сохранение исторической предвзятости в полицейской практике, в то же время сокрытие этой предвзятости за внешним видом или презумпцией механического нейтралитета, известной как «техническая отмывка». Даже если личная или социально-экономическая информация не включена в набор обучающих данных, она, тем не менее, может присутствовать в данных, особенно в отношении места совершения правонарушений. Если в прошлом полиция уделяла непропорционально большое внимание определённому району, то преступления, совершенные в этом районе, были бы раскрыты с большей вероятностью. Это искусственно исказит набор исторических данных, используемый для обучения алгоритма PredPol. В результате PredPol предсказал бы большую вероятность совершения там преступлений в будущем. Тогда больше ресурсов будет направлено на охрану этого района, увековечивая (и, действительно, усиливая) историческую предвзятость, только теперь на якобы «объективной» основе. Если рассматриваемый район был преимущественно населён людьми определённой этнической или религиозной принадлежности - что могло быть той самой причиной, по которой местная полиция исторически сосредоточивала своё внимание именно на нем («этническое профилирование»), – результат мог быть дискриминационным на том основании, что запрещены международным правом прав человека, включая статью 14 Европейской конвенции о правах человека (ETS № 005).

27. Другие полицейские силы Соединённого Королевства в настоящее время также экспериментируют с другой формой прогнозной полицейской деятельности. Девять сил, возглавляемых полицией Уэст-Мидлендса, включая также столичную полицию Лондона и полицию Большого Манчестера, разрабатывают решение для анализа национальных данных (NDAS - см. Ниже). Это будет использовать комбинацию искусственного интеллекта машинного обучения и статистики для оценки, например, риска того, что кто-то совершит или станет жертвой преступления с применением огнестрельного оружия или ножа, а также вероятности того, что кто-то станет жертвой современного рабства. Система основана на данных, касающихся примерно пяти миллионов человек, из которых она определила почти 1400 показателей, которые могут предсказать преступность, из которых 30 являются особенно важными. Полиция Уэст-Мидлендса будет сотрудничать с британским органом по защите данных, чтобы обеспечить соответствие системы правилам конфиденциальности. Офицер, ответственный за проект, признал, что это отчасти является ответом на значительное сокращение бюджетов полиции в последние годы

и вызванное этим давление с целью сосредоточить внимание на лицах, которые нуждаются в срочном вмешательстве.

## **2.1 Прогнозирование повторных правонарушений и предотвращение рецидивов – HART**

28. Система HART была разработана полицией Дарема в сотрудничестве с исследователями из Кембриджского университета с использованием данных обучения 104 000 человек, арестованных в течение пяти лет. В нем используются «прогностические значения», большинство из которых сосредоточены на истории правонарушений подозреваемого, а также на возрасте, поле и географическом регионе, чтобы отнести преступника к категории низкого, среднего или высокого риска совершения новых серьёзных преступлений следующие два года. Лица, относящиеся к категории среднего риска – те, которые «могут совершить несерьёзное преступление» – могут быть затем включены в программу полиции «Контрольно-пропускной пункт», которая «предлагает правомочным правонарушителям четырёхмесячный контракт в качестве альтернативы судебного преследования. В контракте предлагаются меры по устранению основных причин, по которым они совершили преступление, чтобы предотвратить их повторение». HART был разработан с целью сократить количество людей, находящихся в заключении несмотря на то, что они восприимчивы к другим формам вмешательства, которые были бы столь же или даже более эффективны в снижении риска их повторного совершения.

29. Главный полицейский в проекте HART вместе с учёными и австралийским полицейским опубликовал подробную статью, в которой анализируется HART, взяв за основу «принципы необходимости, соразмерности и предсказуемости, изложенные в европейском праве прав человека [в качестве] отправной точки». Примечание. Признавая, что ИИ представляет собой «другой тип принятия решений, а не усовершенствованный человеческий мозг», авторы задают ряд сложных вопросов. «Выдержат ли поэтому судебный контроль и принципы прав человека испытание временем? Какую степень непрозрачности мы готовы принять? Насколько велика ошибка? Насколько велика неопределённость с точки зрения будущих выгод?» Это большие вопросы, актуальные также для каждого применения ИИ в области уголовного правосудия.

30. Сами авторы статьи признают один важный риск или критику, отмеченную также выше в отношении PredPol. «Некоторые из предикторов, используемых в модели (например, postcodeNote), могут рассматриваться как косвенно связанные с показателями депривации сообщества [Один] может возразить, что [использование почтового индекса в качестве] переменной рискует своего рода петлёй обратной

связи, которая может увековечить или усилить существующие модели правонарушений. Если полиция отреагирует на прогнозы, направив свои усилия на районы почтового индекса с наибольшим риском, то больше людей из этих районов попадут в поле зрения полиции и будут арестованы, чем те, кто живёт в районах с меньшим риском, нецелевых. Затем эти аресты становятся результатами, которые используются для создания более поздних версий той же модели, что приводит к постоянно углубляющемуся циклу повышенного внимания полиции». Начальник полиции Дарема Майкл Бартон сказал, что люди должны быть обеспокоены тем, использовался ли протокол HART в судах, но что он использовался для уменьшения числа повторных правонарушений, а не для вынесения приговора. Примечание. При этом не учитывается тот факт, что кто-то неправильно определён как высокорисковый, потому что предвзятости в наборе обучающих данных будет отказано в доступе к программе «Контрольно-пропускной пункт», и вместо этого они могут оказаться под стражей до суда.

31. В 2017 году HART был «пополнен» новыми данными с целью уменьшения зависимости от предикторов почтового индекса. Был использован набор данных под названием «Mosaic», разработанный компанией Experian в первую очередь как коммерческий продукт для маркетинговых целей. Mosaic основан на профилях всех 50 миллионов взрослых жителей Великобритании, собранных с использованием данных, собранных из открытых источников, включая Интернет. Мозаика явно определяет категории, ссылаясь, например, на возрастную группу или этническую принадлежность (например, «разобщённая молодёжь», «зависимые серые» или «азиатское наследие»). Неправительственная организация Big Brother Watch утверждает, что «для компании по проверке кредитоспособности собирать миллионы единиц информации о нас и продавать профили тому, кто предложит самую высокую цену, пугает. Но для полиции подпитывать эти грубые и оскорбительные профили с помощью искусственного интеллекта для принятия решений о свободе и справедливости в Великобритании поистине антиутопия». Примечание. Полиция Дарема заявила, что «работала с Experian, чтобы улучшить своё понимание местных сообществ», Примечание и В 2018 году компания прекратила использование Mosaic, хотя, как сообщается, по финансовым, а не этическим причинам.

32. Хотя полиция Дарема подчеркнула, что протокол HART используется только в консультативных целях и что индивидуальные решения являются обязанностью подготовленных полицейских, некоторые скептически относятся к тому, как все будет работать на практике. Как и в случае использования PredPol полицией Кента, главный констебль Дарема Бартон показал, что неоднократные сокращения бюджета его сил побуждали все чаще прибегать к новым технологиям. Примечание. Эти же сокращения могут иметь последствия для доступности времени и внимания офицеров, что является значительным фактором обеспечения эффективной ответственности человека за решения, принимаемые с использованием

HART. Эндрю Вуфф из Эдинбургского университета заметил, что в «ограниченном по времени и ресурсоёмком» мире полицейской деятельности «я могу представить себе ситуацию, когда полицейский может больше полагаться на систему, чем на свои собственные процессы принятия решений». Примечание. Компания Big Brother Watch утверждала, что «с учётом того, что алгоритм был разработан для выявления случаев, которые полиция может пропустить или не склонна считать высокорисковыми, действительно ли возможно ожидать, что полицейские будут последовательно выносить приговоры в отношении ИИ результат? Чтобы действовать как средство принятия решений, он должен предупреждать полицию о потенциальных нарушителях, которых они могли не учитывать. Поэтому сомнительно, что полиция спокойно проигнорирует эти предложения».

33. Big Brother Watch также привлекли внимание к вопросам объяснимости и подотчётности, отметив, что «решения ИИ не могут быть оспорены, потому что может быть невозможно даже объяснить сделанные выводы». Полиция Дарема ответила, что «будет готова раскрыть алгоритм HART и связанные с ним персональные данные и наборы данных событий хранения алгоритмическому регулятору». Примечание. Это не необоснованная позиция, но в то же время сильный аргумент в пользу создания таких тела.

## **2.2 Прогнозирование повторного совершения преступления и решение о предварительном заключении, приговоре и условно-досрочном освобождении – COMPAS**

34. COMPAS, система, в настоящее время принадлежащая компании Equivant, используется в нескольких юрисдикциях штатов США для оценки риска повторного совершения преступления. Нортпойнт, ныне подразделение Equivant, охарактеризовал его как «инструмент оценки рисков и потребностей четвёртого поколения». Это веб-инструмент, предназначенный для оценки криминогенных потребностей преступников и риска рецидива» с использованием трёх «шкал»: «риск досудебного освобождения» (т.е. риск неявки и ареста за новое уголовное преступление); «Общий рецидив» (совершение нового мисдиминора или уголовного преступления в течение двух лет); и «насильственный рецидив» (совершение насильственных преступлений). Отметим, что, например, штат Нью-Йорк начал с использования системы COMPAS для оценки людей, отбывающих испытательный срок; теперь он используется там также судьями при вынесении приговоров. Во Флориде он используется при принятии решения о заключении обвиняемого под стражу или под залог. В Висконсине он используется на каждом этапе тюремной системы, от вынесения приговора до условно-досрочного освобождения.

35. В 2014 году тогдашний генеральный прокурор США Эрик Холдер предупредил, что «хотя [оценки рисков типа COMPAS] были разработаны с наилучшими намерениями, я обеспокоен тем, что они непреднамеренно подрывают наши усилия по обеспечению индивидуального и равного правосудия. Они могут усугубить необоснованное и обостренное, несправедливое неравенство, которое уже слишком распространено в нашей системе уголовного правосудия и в нашем обществе». В ответ на эти опасения исследовательский вебсайт ProPublica изучил использование COMPAS и пришел к двум особенно важным выводам. Что касается его эффективности, ProPublica обнаружила, что «когда принимается во внимание весь спектр преступлений алгоритм был несколько более точным, чем подбрасывание монеты», и он был «чрезвычайно ненадежным при прогнозировании насильственных преступлений». Что касается нейтралитета, ProPublica обнаружила, что, хотя COMPAS делал ошибки примерно с одинаковой скоростью как для белых, так и для чернокожих, он с гораздо большей вероятностью давал ложноположительные результаты (а именно, ошибочный прогноз «высокого риска») для чернокожих и с большей вероятностью ложноотрицательные результаты (а именно ошибочное предсказание низкого риска) для белых людей.

36. Следует отметить, что выводы ProPublica подвергались критике как по техническим причинам, связанным с достоверностью статистического анализа, так и по поводу того, как были представлены результаты. *Примечание. Однако даже автор одного из этих критических опровержений предположил, «что похоже предвзятость не в инструменте, а в системе». Примечание. Это вновь вводит понятие «технологической промывки» – и разрушительные, дискриминационные последствия для людей во многом те же, независимо от описания проблемы.*

37. Споры по поводу COMPAS также побудили комментаторов подчеркнуть необходимость публичных дебатов, прозрачности и подотчётности. «Демократические общества должны сейчас работать, чтобы определить, какой прозрачности они ожидают от систем ADM [автоматизированного принятия решений]. Нужны ли нам новые правила использования программного обеспечения, чтобы его можно было должным образом проверить? Законодатели, судьи и общественность должны иметь право голоса при выборе алгоритмов приоритета мер справедливости. Но если алгоритмы на самом деле не отражают эти оценочные суждения, кто будет нести ответственность?» *Примечание. Некоторые утверждали, что такая прозрачность должна включать доступ к алгоритму: «Сложные проблемы, возникающие при использовании программного обеспечения, такого как [COMPAS], являются веской причиной сделать формулы общедоступными или, по крайней мере, подвергнуть их тщательной независимой проверке».*

### 2.3 Выявление разрешимых исторических случаев - «Календарь холодных дел»

38. Полиция Нидерландов разработала систему машинного обучения на основе ИИ, чтобы помочь им выявлять старые, нераскрытые, серьёзные дела («нераскрытые дела»), которые теперь могут иметь хорошие перспективы для решения. Эта система основана на понимании того, что более чем в половине случаев вновь открытые холодные дела решаются благодаря новой технологии, которая не была доступна во время первого расследования. (Почти в половине случаев это происходит из-за новых советов свидетелей; термин «календарь незавершённых дел» первоначально относился к методу опроса сокамерников по «незавершённым делам».) После оцифровки досье «незавершённых дел» они вводятся в систему искусственного интеллекта, которая определяет те, которые содержат многообещающие доказательства, которые могут быть повторно исследованы с использованием новых методов судебной экспертизы. Подобная работа, выполняемая сотрудниками полиции вручную, могла занять несколько недель на одно дело несмотря на то, что вероятность успеха была отдалённой. Офицеры, ответственные за проект, надеются, что его можно будет расширить для выявления «холодных дел», которые могут быть раскрыты с использованием данных, не связанных с судебной медициной, таких как социальные науки, социальные сети и показания свидетелей. Это могло бы даже оказаться способным улучшить способность полиции вести расследование правонарушений.

39. Одна проблема с этим подходом заключается в том, что для полицейских данных, которые не использовались в первоначальном расследовании, установленные законом сроки хранения были слишком короткими для расследований по нераскрытым делам. Таким образом, уничтожение таких данных могло бы устранить возможную основу для раскрытия дела с помощью новых методов судебной экспертизы. Поэтому начальник полиции решил не уничтожать такие старые данные. Вместо этого теперь доступ к нему ограничен, и доступ к нему имеет только ограниченное число так называемых «привратников». Правительство приняло это решение, заявив, что «было бы предпочтительнее принять этот недостаток в соответствии с законом и согласиться с шагами, предпринятыми начальником полиции для ограничения доступа к данным до строго необходимых. Соблюдение буквы закона может быть достигнуто только с помощью метода грубого отбора, который также уничтожит данные, которые могут способствовать обнаружению в холодных случаях. Это серьёзно помешало бы разрешению этих дел. В настоящее время это предотвращается». Примечание. Однако неясно, соответствует ли этот подход общему принципу закона о защите данных, согласно которому персональные данные должны обрабатываться (включая сохранение) только в том случае, если есть основание в законе или согласие субъекта данных.

## 2.4 Другие примеры

40. Не все алгоритмические системы, предназначенные для помощи в принятии решений, включают искусственный интеллект. Исследовательский проект в США показал, что суды в 46 из 52 штатов США, а также в округе Колумбия, используют тот или иной инструмент оценки рисков в процессе принятия решений о предварительном заключении. Примечание. Большинство из них «не обучаются» алгоритмическим инструментам». Примечание. Наиболее распространёнными инструментами были оценка общественной безопасности (PSA), используемая как минимум в пяти штатах и 59 округах (административных единицах штатов), охватывающая 56,3 миллиона человек; Инструмент досудебной оценки рисков Вирджинии (VPRAI), используемый по крайней мере в 43 округах с охватом 19,9 миллионов человек, и его пересмотренная версия (VPRAI-R, предположительно исправленная с учётом расовой и гендерной предвзятости), используемая как минимум в одном штате и 16 округах с охватом 14,3 миллиона человек. люди; Инструмент предварительной оценки системы оценки рисков штата Огайо (ORAS-PAT), используемый по крайней мере в пяти штатах и 48 округах; и COMPAS (см. выше), используемый как минимум в 11 округах с населением 4,3 миллиона человек.

41. Институт досудебного правосудия (PJI), американская некоммерческая организация, в прошлом продвигал использование инструментов алгоритмической оценки риска в качестве средства уменьшения количества обращений за наличным залогом (условное освобождение при выплате определённой суммы денег), которое часто приводило к заключению под стражу менее состоятельных обвиняемых. Однако в июле 2020 года он изменил свою позицию, заметив, что, несмотря на снижение показателей содержания под стражей до суда, этнический состав задержанных оставался примерно на 50% чёрным и 30% белым. В объявлении о новой политике говорилось, что: «Теперь мы видим, что инструменты досудебной оценки рисков, предназначенные для прогнозирования явки человека в суд без нового ареста, больше не могут быть частью нашего решения по построению справедливых систем досудебного правосудия. Независимо от их науки, бренда или возраста, эти инструменты основаны на данных, отражающих структурный расизм и институциональное неравенство, которые влияют на нашу судебную и правоохранительную политику и практику. Использование этих данных усугубляет неравенство».

42. Критика исходила также из органов государственного управления. В отчёте общего суда штата Массачусетс говорится, что «какими бы благими намерениями ни были инструменты оценки риска, они могут иметь свои ограничения из-за их зависимости от данных сомнительной корреляции с предсказуемостью и жёсткостью применения, которая ограничивает решение судьи на исторических данных для обработки и определения вероятного результата

применительно к отдельному ответчику. Качество прогноза во многом зависит от качества данных, вводимых в инструмент. Существенный недостаток многих инструментов оценки риска состоит в том, что алгоритм полагается на данные об аресте в качестве отправной точки анализа для прогнозирования будущего события, и эти данные могут включать неявную систематическую ошибку. Более того, многие коммерчески доступные инструменты оценки риска не раскрывают свои алгоритмы или методологии, которая порождает недоверие к системе, затрудняет оспаривание защитником своих результатов и ограничивает способность юрисдикции быстро и точно решать противоречивые результаты. Опора на данные об арестах как на фактор прогнозирования привела к смещённым результатам, контрпродуктивным для цели использования инструмента в первую очередь. Недостатки внедрения доступного в настоящее время инструмента оценки рисков, вероятно, перевесят любое постепенное улучшение решений об освобождении под залог».

43. Судьи Верховного суда штата Огайо приняли решение не рекомендовать использование инструментов досудебной оценки рисков в рамках реформы государственной системы освобождения под залог, несмотря на предыдущие выводы специальной целевой группы. Утверждается, что на судей большое влияние оказали аргументы, представленные Американским союзом гражданских свобод, в которых особо подчёркивался риск расовой предвзятости.

### **3. Надзор и регулирование**

#### **Комитет по этике полиции Уэст-Мидлендса и комиссара по преступности Великобритании**

44. В этом разделе я сосредоточусь на ситуации с полицией Уэст-Мидлендс (WMP), которая хорошо развита и о которой имеется значительный объём информации (в том числе через мои собственные контакты с заинтересованными лицами). WMP имеет проект «Data-Driven Insights», предназначенный для более эффективного использования больших объёмов данных, уже содержащихся в его различных наборах данных, которые существуют как отдельные информационные «хранилища», содержимое которых нелегко объединить или связать перекрёстными ссылками. Понимая, что те же инструменты обработки данных можно легко передать другим 43 полицейским силам Соединённого Королевства, этот проект был расширен до национального решения для анализа данных (NDAS). NDAS разрабатывается в сотрудничестве с восемью другими полицейскими силами или агентствами Великобритании и при технической поддержке частной компании Accenture. Проект состоит из трёх основных элементов: Insight Search, поисковая

машина, охватывающая девять баз данных, доступная офицерам на мобильных устройствах; Business Insights, который фокусируется на внутренних вопросах (например, благополучии сотрудников); и Insights Lab, которая разрабатывает инновационные инструменты анализа данных специально для работы полиции. Бюджет проекта Data-Driven Insights составлял 17 миллионов фунтов стерлингов, при этом Insight Search, по оценкам, принёс выгоды на сумму 23 миллиона фунтов стерлингов только за первые три года работы. До сих пор его системы на основе искусственного интеллекта в основном использовались для выявления лиц, которые могут продолжать совершать преступления с применением ножа или огнестрельного оружия («Наиболее серьёзное насилие»), а также для выявления лиц, которые могут стать жертвами торговли людьми («Современное рабство»).

45. Комиссар полиции и преступности Уэст-Мидлендса (РСС) является избранным независимым должностным лицом, которому поручено обеспечивать эффективность и действенность WMP и подотчётность его главного констебля. РСС Уэст-Мидлендса учредил комитет по этике для консультирования РСС и главного констебля по проектам в области науки о данных, предлагаемым лабораторией аналитики данных WMP. Заявленная цель этого комитета – «обеспечить, чтобы этика и права людей были в центре работы лаборатории». Его круг ведения требует, чтобы он учитывал широкий спектр этических принципов; несмотря на первоначальную сдержанность со стороны WMP, есть прямая ссылка на права человека, как в целом, так и более конкретно в отношении недискриминации и неприкосновенности частной жизни. Примечание. Комитет намеренно разнообразен, с гендерным балансом 50:50 и наймом, которым занимается агентство, специализирующееся на этническом разнообразии, которому поручено доводить вакансии до сведения «этнических меньшинств, родителей-одиночек, инвалидов, религиозных, сексуальных и гендерных групп». Большинство участников имеют определённую техническую подготовку, обратите внимание, хотя есть и неспециалисты. В состав комитета входит высокопоставленный представитель другой силы, которая, как сказал мне Макнил, помогает «обосновать комитет на самом деле». Его деятельность в высшей степени прозрачна и включает работу с населением и консультации, хотя и в рамках ограниченного бюджета. Было решено, что, хотя NDAS является результатом сотрудничества нескольких полицейских сил, этический комитет РСС Вест-Мидлендса также будет консультировать по предложениям NDAS, поскольку альтернативной национальной этической структуры не существует. Тем не менее, одним из потенциальных структурных недостатков Комитета по этике является то, что его существование зависит от РСС Вест-Мидлендса; нет никакой гарантии, что любой будущий РСС будет поддерживать его.

46. Делая предложения в Комитет по этике, Лаборатория аналитики данных использует матрицу ALGO-CARE. Он был разработан г-жой Освальд и другими после обзора программы HART полиции Дарема (см. Выше) и был принят

Национальным советом начальников полиции в качестве модели для передовой практики саморегулирования. Он призван побудить разработчиков алгоритмических инструментов для использования полицией рассмотреть ряд существенных практических и этических проблем. ALGO-CARE – это аббревиатура, используемая для объединения ряда вопросов, касающихся этих проблем, включая следующие:

Консультативный (используется ли алгоритмический вывод в качестве консультативного, когда человек отвечает за принятие решений?);

Законный (преследует ли алгоритм законную цель, соразмерен ли он и были ли использованные данные получены и обработаны законным образом?);

Гранулярность (достаточно ли подробны предложения алгоритма, надёжны ли входные данные и исправлены ли возможные ошибки?);

Право собственности (кому принадлежит алгоритм и данные, есть ли у полиции все необходимые права для их использования, и будет ли система поддерживаться, обновляться и защищаться по мере необходимости?);

Спорный (каковы механизмы надзора и аудита? Уведомляются ли субъекты и информируются о его использовании?);

Точность (соответствует ли указанный уровень точности цели политики и можно ли его периодически проверять? Можно ли обосновать частоту ложных срабатываний/отрицательных результатов, каковы последствия ложных результатов и приемлем ли результирующий риск? инструмент есть необходимая экспертиза?);

Ответственное (является ли использование инструмента объективно справедливым, прозрачным и подотчётным? Будет ли это рассмотрено в общественных интересах и этично?);

Объяснимый (доступна ли соответствующая информация о правилах и оценке различных факторов? Сможет ли полиция попросить эксперта по науке о данных объяснить и обосновать инструмент?).

47. Комитет по этике не выносит обязательных решений, а скорее даёт рекомендации РСС и главному констеблю. Совет касается не закона или соблюдения правовых норм, а в более широком смысле этики. В принципе, главный констебль может проигнорировать этот совет, но он / она останется подотчётен РСС, который теоретически может уволить его / её. На практике советам комитета всегда следовали. Например, предложение о приложении «интегрированного управления правонарушителями» прошло три итерации, прежде чем получить одобрение Комитета по этике. Г-н Макнейл сказал мне, что ряд предложений был отклонён, в том числе модель прогнозирования, полученная после отзывов от различных общественных организаций.

48. Из моего разговора со старшим суперинтендантом Годдом и детективом старшим инспектором Дейлом было очевидно, что они хорошо понимали этические вопросы, связанные с NDAS. Мы обсудили вопросы защиты данных, общественного доверия, прозрачности, объяснимости, подотчётности и ответственности, и надзора; а также необходимость технологических решений для доказательства их эффективности, которые необходимо будет оценить, чтобы определить соразмерность любого вмешательства в защищаемые права. Офицеры подчеркнули, что всегда будет «человек в петле», который будет принимать (и брать на себя) окончательные решения – алгоритмический вывод будет только информировать, а не определять решение. Они также описали, как потенциальные проблемы исторической предвзятости в наборах данных поднимались и решались во время взаимодействия между полицией и Комитетом по этике.

49. Главный Супт. Годд и DCI Дейл знали, как и г-н Макнил (и г-жа Освальд, во время слушаний в нашей комиссии), что Комитет по этике не был идеальным решением. Они также подчеркнули тот факт, что в Великобритании нет специального законодательства об использовании ИИ полицией. Колледж полицейской службы в настоящее время разрабатывает авторизованную профессиональную практику, общий национальный стандарт по использованию аналитики данных, но это тоже не будет обязательным; однако несоблюдение этого требования может быть важным соображением в ходе судебного разбирательства.

### **3.1 Элементы возможного регулирования ИИ, используемого в системах охраны правопорядка и уголовного правосудия**

50. Д-р Виль и г-жа Освальд также внесли предложения о том, как следует регулировать ИИ. Использование алгоритмических систем в процессах принятия решений людьми должно быть обосновано в каждом конкретном случае с учётом операционного контекста. Это должно включать рассмотрение характера результирующих вмешательств и их воздействия, особенно на маргинализированные сообщества. В государственном секторе должен быть национальный реестр систем ИИ. Особое внимание следует уделять кумулятивному эффекту алгоритмических систем: одна, по отдельности, может иметь только пропорциональные неблагоприятные эффекты, но несколько систем, используемых на разных этапах обработки данных, могут иметь гораздо более сильные, трудно прогнозируемые кумулятивные воздействия. Системы должны быть понятны их непосредственным пользователям и тем, на кого влияет процесс принятия решений. Должна быть возможность тщательно изучить и протестировать систему ИИ на месте эксплуатации. Проприетарные системы, не допускающие этого, следует исключить. Следует тщательно изучить последствия внедрения элементов ИИ в

существующие технологии: например, ИИ с распознаванием лиц, применяемый к сетям видеонаблюдения, уже создал чрезвычайно мощную форму массового наблюдения, выходящую далеко за рамки первоначальных ожиданий, когда система видеонаблюдения была впервые введена, и также имела потенциал для автоматического чтения по губам.

### **3.2 САНАИ и возможная правовая база для искусственного интеллекта**

51. В сентябре 2019 года Комитет министров учредил Межправительственный специальный комитет по искусственному интеллекту (САНАИ). САНАИ было поручено изучить возможность и потенциальные элементы правовой базы для разработки, проектирования и применения искусственного интеллекта. Его работа основана на стандартах Совета Европы в области демократии, прав человека и верховенства закона, а также на других соответствующих международно-правовых документах и на текущей работе других международных и региональных организаций. Помимо обычных участников, представляющих государства — члены Совета Европы, государства — наблюдатели и другие органы Совета Европы (включая Ассамблею), САНАИ имеет исключительно высокий уровень взаимодействия с органами частного сектора, гражданским обществом, исследовательскими и академическими учреждениями.

52. САНАИ провела своё первое заседание 18-20 ноября 2019 года. Среди прочего, было решено, что ключевым элементом будущего технико-экономического обоснования будет «отображение рисков и возможностей, возникающих в результате проектирования, разработки и применения искусственного интеллекта включая влияние последнего на права человека, верховенство закона и демократию». В настоящее время САНАИ планирует принять технико-экономическое обоснование на своём третьем заседании, запланированном на декабрь 2020 года.

53. Это институциональный контекст, в котором Ассамблея будет обсуждать настоящий и различные другие отчёты, связанные с ИИ, подготовленные в различных комитетах. Ассамблея решила подойти к этой теме на основе контекста, изучив влияние ИИ в различных областях. В Комитете по правовым вопросам и правам человека, например, также есть отчёты о технологиях интерфейса мозг-компьютер, автономных транспортных средствах и (на ранних стадиях подготовки) летальных автономных системах вооружений. Таким образом, рекомендации, которые Ассамблея может принять на основе этих отчётов, послужат важным руководством для САНАИ при составлении карты рисков и возможностей ИИ и его воздействия на права человека, верховенство закона и демократию и последующем определении необходимости обязательного международно-правовая база.

54. Использование инструментов искусственного интеллекта в системах охраны правопорядка и уголовного правосудия представляет особые риски для прав человека. С одной стороны, это связано с важностью решений, которые могут быть приняты на основе алгоритмического вывода - решений о слежке, обысках и изъятии, аресте, задержании, вынесении приговора, освобождении под залог или условно-досрочное освобождение и т. д. Все из которых имеют особенно сильное воздействие на права человека. С другой стороны, это связано с тем, что исторические полицейские данные, используемые ИИ для обучения алгоритмов, часто испорчены историческими предубеждениями, а также потому, что внедрение ИИ в такой деликатный процесс принятия решений может подорвать человеческую ответственность и подотчётность и сделать решения недействительными прозрачный и сложный. Мои обсуждения с полицией Уэст-Мидлендса в Соединённом Королевстве показали, что полиция сама осознает необходимость правового регулирования использования ИИ в полицейской деятельности и в отсутствие правового регулирования уже предпринимает собственные шаги для обеспечения этического надзора и стандартизированной передовой практика. Строгие общие требования законности и надлежащей правовой процедуры, изложенные в статьях 5 и 6 Конвенции – права на свободу и безопасность и на справедливое судебное разбирательство, которые специально предназначены для применения к полиции и системе уголовного правосудия подтверждают необходимость в правовом регулировании.

55. На уровне международных стандартов применимые нормы уже существуют, особенно в Европейской конвенции о правах человека и прецедентном праве Европейского суда по правам человека, а также в специализированных инструментах, таких как Конвенция 108. Однако в связи с характером искусственного интеллекта *suī generis* и новыми способами его применения соответствующие стандарты лучше всего были бы дополнительно определены в специальном обязательном документе.

## **Выводы и рекомендации**

56. Хотя исследования в области ИИ на базе цифровых компьютеров начались ещё в середине прошлого века, его разработка и применение в последние годы значительно продвинулись вперёд, во многом благодаря достижениям в методах машинного обучения, которые стали возможными благодаря увеличению компьютерной обработки, мощности и доступности массивных наборов обучающих данных. В настоящее время ИИ используется в широком и постоянно расширяющемся диапазоне ситуаций, некоторые из которых могут иметь серьёзные

последствия для демократии, прав человека и верховенства закона, включая систему уголовного правосудия.

57. В этом отчёте исследованы некоторые общие вопросы и проблемы, связанные с ИИ, и рассмотрены четыре конкретных применения ИИ в различных аспектах системы уголовного правосудия. Он также рассмотрел некоторые из конкретных проблем, поднятых в отношении определённых приложений, а также систему этического надзора за использованием ИИ полицией. Мой анализ показывает, что общие опасения по поводу ИИ в равной степени применимы к конкретным приложениям ИИ в полицейских системах и системах уголовного правосудия, которые я изучал. Это также предполагает, что общие аргументы в пользу более строгого регулирования ИИ могут иметь особое отношение к его применению в системах уголовного правосудия. Мои подробные предложения по реагированию на эту ситуацию изложены в прилагаемом проекте резолюции и рекомендации.

## **Искусственный интеллект – описание и этические принципы**

Было много попыток дать определение термину «искусственный интеллект» с тех пор, как он впервые был использован в 1955 году. Эти усилия усиливаются по мере того, как органы, устанавливающие стандарты, включая Совет Европы, реагируют на растущую мощь и повсеместность искусственного интеллекта, работая в направлении его правовое регулирование. Тем не менее, до сих пор не существует единого общепринятого «технического» или «юридического» определения. Однако для целей настоящего отчёта необходимо описать концепцию.

Термин «искусственный интеллект» (ИИ) в настоящее время обычно используется для описания компьютерных систем, которые могут воспринимать и извлекать данные из своей среды, а затем использовать статистические алгоритмы для обработки этих данных с целью получения результатов, предназначенных для достижения заранее определённых целей. Алгоритмы состоят из правил, которые могут быть установлены человеком или заданы самим компьютером, который «тренирует» алгоритм, анализируя массивные наборы данных, и продолжает уточнять правила по мере получения новых данных. Последний подход известен как «машинное обучение» (или «статистическое обучение») и в настоящее время является наиболее широко используемым методом для сложных приложений, который стал возможен только в последние годы благодаря увеличению вычислительной мощности компьютера и доступности достаточного количества данных. «Глубокое обучение» – это особенно продвинутая форма машинного обучения, использующая несколько уровней «искусственных нейронных сетей» для обработки данных. Алгоритмы, разработанные этими системами, могут быть не

полностью восприимчивы к человеческому анализу или пониманию, поэтому их иногда называют «чёрными ящиками» (термин, который также, но по другой причине, иногда используется для описания закрытых систем ИИ, защищённых правами интеллектуальной собственности).

Все текущие формы ИИ «узкие», то есть они предназначены для решения одной определённой задачи. «Узкий» ИИ также иногда называют «слабым», даже если современные системы распознавания лиц, обработки естественного языка, автономного вождения и системы медицинской диагностики, например, невероятно сложны и выполняют определённые сложные задачи с поразительной скоростью и точностью. «Общий искусственный интеллект», иногда называемый «сильным» ИИ, способный выполнять все функции человеческого мозга, все ещё находится в будущем. «Искусственный супер-интеллект» относится к системе, возможности которой превышают возможности человеческого мозга.

По мере того, как количество областей, в которых применяются системы искусственного интеллекта, растёт, распространяясь на области со значительным потенциальным воздействием на индивидуальные права и свободы, а также на системы демократии и верховенства закона, все большее и все более безотлагательное внимание уделяется этическому аспекту.

Широкий круг участников выдвинул множество предложений по набору этических принципов, которые должны применяться к системам ИИ. Эти предложения редко бывают идентичными, различаются как принципами, которые они включают, так и способами определения этих принципов. Исследования показали, что, тем не менее, существует широкое согласие по основному содержанию этических принципов, которые должны применяться к системам искусственного интеллекта, а именно по следующим:

*Прозрачность.* Принцип прозрачности можно широко интерпретировать, включая доступность и объяснимость системы ИИ, другими словами, возможности для человека понять, как система работает и как она даёт свои результаты.

*Справедливость и честность.* Этот принцип включает недискриминацию, беспристрастность, последовательность и уважение разнообразия и плюрализма. Это также подразумевает возможность для субъекта работы системы ИИ оспаривать результаты с возможностью исправления и возмещения.

*Обязанность.* Этот принцип включает в себя требование о том, что человек должен нести ответственность за любое решение, затрагивающее права и свободы человека, с определённой ответственностью и юридической ответственностью за эти решения. Таким образом, этот принцип тесно связан с принципом справедливости и беспристрастности.

*Безопасность и охрана.* Это означает, что системы ИИ должны быть надёжными, защищёнными от внешнего вмешательства и безопасными от выполнения непреднамеренных действий в соответствии с принципом предосторожности.

*Конфиденциальность.* Хотя уважение прав человека в целом можно рассматривать как неотъемлемую часть принципов справедливости и беспристрастности, а также безопасности и защищенности, право на неприкосновенность частной жизни особенно важно там, где система искусственного интеллекта обрабатывает личные или частные данные. Поэтому системы искусственного интеллекта должны соблюдать обязательные стандарты Общего регламента ЕС по защите данных (GDPR) и Конвенции Совета Европы о защите частных лиц в отношении автоматической обработки персональных данных (ETS No. 108 и «модернизированная» конвенция 108+, CETS No. 223), если применимо.

Эффективное внедрение этических принципов в отношении систем искусственного интеллекта требует подхода, основанного на принципе «этики», включая оценку воздействия на права человека, чтобы гарантировать соблюдение установленных стандартов. Недостаточно разрабатывать системы только на основе технических стандартов и добавлять элементы на более поздних этапах в попытке продемонстрировать уважение к этическим принципам.

Степень, в которой соблюдение этих принципов должно быть встроено в конкретные системы искусственного интеллекта, зависит от предполагаемого и прогнозируемого использования этих систем: чем больше потенциальное воздействие на общественные интересы, права и свободы личности, тем более строгие меры защиты нужны. Таким образом, этическое регулирование может осуществляться различными способами: от добровольных внутренних уставов для наименее уязвимых областей до обязательных правовых стандартов для наиболее чувствительных. Во всех случаях он должен включать независимые механизмы надзора в зависимости от уровня регулирования.

Эти основные принципы сосредоточены на системе ИИ и её непосредственном контексте. Они не предназначены для того, чтобы быть исчерпывающими или исключать более широкие этические проблемы, такие как демократия (плюралистическое участие общественности в разработке этических и нормативных стандартов), солидарность (признание различных точек зрения различных групп) или устойчивость (сохранение планетарной среды).

## **Интерфейс мозг-компьютер: новые права или новые угрозы основным свободам?**

Комитет: Комитет по правовым вопросам и правам человека

Докладчик: Г-н Оливье БЕХТ, Франция, ALDE

Происхождение: Обращение в комитет: Док. 14814, номер 4435 от 12 апреля 2019 г. 2020 г. – октябрь Постоянный комитет

### **Краткое содержание**

Быстрый прогресс, достигнутый в нейротехнологии в последние годы, может создать всё более эффективные интерфейсы мозг-компьютер (BCI). Доступ к нейронным процессам, лежащим в основе сознательного мышления, подразумевает доступ к уровню «я», который по определению не может быть сознательно скрыт или отфильтрован. Это чревато серьёзным нарушением частной жизни и достоинства личности, что может подрвать свободу воли и нарушить высшее убежище человеческой свободы – разум. Когнитивные и сенсорные улучшения с помощью BCI могут создать отдельные категории людей. Индивидуальная идентичность, субъектность и моральная ответственность могут быть уменьшены за счёт слияния неврологического и цифрового сенсорного опыта и процессов принятия решений. Такие результаты могут изменить саму природу человечества и человеческих обществ.

Демократические общества должны гарантировать соблюдение основных этических принципов. Огромные потенциальные преимущества нейротехнологий таковы, что нельзя сдерживать прогресс и инновации. Тем не менее, исследования следует направлять в сторону от предсказуемо вредных или опасных областей в сторону положительных применений.

Комитет считает, что необходим чувствительный, откалиброванный подход к регулированию появляющихся нейротехнологий, включая технологию BCI, охватывающий как этические рамки, так и обязательное правовое регулирование. Поэтому он призывает государства-члены, соответствующие межправительственные комитеты Совета Европы и Комитет министров предпринять конкретные шаги с этой целью.

## А. Проект резолюции

1. Парламентская ассамблея отмечает быстрый прогресс, достигнутый в области нейротехнологий за последние годы, включая возможность записывать и напрямую стимулировать нейронную активность, что даёт возможность создавать все более эффективные интерфейсы мозг-компьютер (BCI). Этот прогресс был обусловлен сочетанием улучшенного понимания функционирования мозга, технических достижений и растущей мощности систем искусственного интеллекта. Возможность создать полностью симбиотическую связь между человеческим мозгом и цифровыми вычислительными системами, включая Интернет и системы искусственного интеллекта, остаётся далёкой целью. Тем не менее, это цель, которую исследователи и предприниматели уже преследуют и которая, по мнению многих, в конечном итоге может быть достигнута.

2. Нейротехнологии, в том числе BCI, в настоящее время разрабатываются и применяются для различных целей. Среди прочего, огромные суммы вкладываются в исследования по созданию новых методов лечения неврологических и психических расстройств, таких как прямой контроль над роботизированными конечностями, синтез речи или лечение трудноизлечимых расстройств настроения или посттравматического стрессового расстройства. Военные учреждения и службы безопасности изучают нейротехнологию для использования в разведке, пропаганде, допросах, наблюдении и повышении эффективности комбатантов. Частные компании изучают возможность использования потребительских устройств для преобразования мыслей непосредственно в набор текста; предоставление коммерческих услуг по обнаружению лжи на основе сканирования мозга; и продажа нейротехнологических устройств непосредственно потребителю, например, в качестве продуктов для компьютерных игр или оздоровления. Исследователи изучают возможности разработки «нейромаркетинговых» кампаний, которые будут использовать подсознательные предпочтения, и изучают, могут ли модели нейронной активности предсказывать рецидив преступников.

3. Доступ к нейронным процессам, лежащим в основе сознательного мышления, подразумевает доступ к уровню «я», который по определению не может быть сознательно скрыт или отфильтрован. Это чревато серьёзным нарушением частной жизни и достоинства личности, что может подорвать свободу воли и нарушить высшее убежище человеческой свободы – разум. Когнитивные и сенсорные улучшения с помощью BCI могут создать отдельные категории людей, улучшенные и неулучшенные, с улучшением, доступным только тем, у кого есть необходимое богатство и привилегии, или используемыми в репрессивных целях. Индивидуальная идентичность, субъектность и моральная ответственность могут быть уменьшены за счёт слияния неврологического и цифрового сенсорного опыта

и процессов принятия решений. Такие результаты могут изменить саму природу человечества и человеческих обществ.

4. Даже если наиболее впечатляющие гипотетические применения ВСІ остаются спекулятивными, уже достигнутые успехи и ресурсы, выделяемые на дальнейшие исследования, подразумевают острую необходимость в упреждающем и предупредительном регулировании сейчас. Демократические общества должны гарантировать соблюдение основных этических принципов. Огромные потенциальные преимущества нейротехнологий, особенно в области медицины, таковы, что прогресс и инновации не должны подавляться. Тем не менее исследования следует направлять в сторону от предсказуемо вредных или опасных областей и в сторону позитивных приложений, которые не угрожают достоинству, равенству и свободе личности, которые также являются основами демократии.

5. Ассамблея считает, что необходим чувствительный, выверенный подход к регулированию возникающих нейротехнологий, включая технологию ВСІ, охватывающий как этические рамки, так и обязательное правовое регулирование. В нем отмечаются сходства и связи между «нейроэтикой» и биоэтикой, а также значение искусственного интеллекта для работы технологии ВСІ. Поэтому он приветствует работу, уже ведущуюся в рамках Совета Европы Комитетом по биоэтике (DH-BIO) и Специальным комитетом по искусственному интеллекту (САНАИ). Он также приветствует работу других международных организаций, в частности Организации экономического сотрудничества и развития (ОЭСР), которая недавно приняла Рекомендацию об ответственных инновациях в нейротехнологиях. Ассамблея с интересом отмечает развитие событий, например, в Чили, где рассматривается возможность внесения поправок в конституцию, законодательства и других мер, направленных на защиту человеческого общества от возможных неблагоприятных последствий нейротехнологий.

6. Ассамблея считает, что следующие этические принципы должны применяться к развитию и применению нейротехнологии в целом и технологии ВСІ в частности:

6.1. Польза и предотвращение злонамеренного использования. Эта технология должна разрабатываться и применяться только в целях, соответствующих уважению прав и достоинства человека. Следует запретить исследования, направленные на несовместимые цели. Особое внимание следует уделять технологиям двойного назначения и технологиям, разработанным для военных целей или безопасности. Новые нейротехнологии должны быть подвергнуты предварительной оценке воздействия на права человека перед вводом в эксплуатацию.

6.2. Безопасность и меры предосторожности. Эта технология должна быть безопасной как для пользователя, так и для общества в целом с учётом

предполагаемых или непредвиденных последствий. Перед началом использования любых новых приложений необходимо обеспечить безопасность.

6.3. Секретность и конфиденциальность. Как минимум, информация, собираемая нейротехнологическими устройствами и устройствами ВСІ, должна быть защищена в соответствии с общими принципами защиты данных. Следует также рассмотреть возможность защиты «нейроданных» как особой категории, например, по аналогии с запретом на торговлю человеческими органами.

6.4. Ёмкость и автономность. Эта технология не должна использоваться против воли субъекта или таким образом, чтобы он не мог свободно принимать дальнейшие решения относительно их дальнейшего использования. Особое внимание потребуется там, где такая технология используется для лечения хронической боли, лекарственной зависимости или других состояний, при которых прерывание лечения может привести к дискомфорту или стрессу.

6.5. Человеческая деятельность и ответственность. Эта технология не должна мешать человеку работать свободно и нести ответственность за свои действия. Люди, действующие свободно в соответствии со своим естественным (в отличие от расширенного или симбиотического) сознания, должны оставаться единственными лицами, принимающими решения, и основными действующими лицами в обществе, особенно в вопросах, которые могут повлиять на права человека и демократические процессы.

6.6. Справедливость, честность и открытость. Эта технология не должна создавать какие-либо привилегии или высший статус для их пользователей; он должен осуществляться с уважением человеческого равенства и достоинства, в том числе членов маргинализированных или уязвимых групп; и они должны быть доступны как можно шире, особенно если они применяются в медицинских целях.

6.7. Обеспечение общественного доверия через прозрачность, консультации и просвещение / повышение осведомлённости. Внедрение новых технологий, таких как нейротехнологии, предназначенных для использования отдельными людьми, будет лучше всего приветствоваться и приниматься, если оно будет происходить с доверием общественности, с осознанием преимуществ, а также потенциальных опасностей.

7. Степень, в которой технология ВСІ может иметь потенциал для фундаментального изменения отношений между внутренним и подсознательным «я» человека и внешним миром, предполагает уникальные и беспрецедентные угрозы фундаментальным ценностям прав и достоинства человека. Ассамблея с особым интересом отмечает предложения по установлению и обеспечению правовой защиты новых прав человека, иногда называемых «неврологическими». Эти предложения призваны заполнить пробелы в существующей системе прав человека, через которые технология ВСІ может угрожать осуществлению

защищаемых в настоящее время прав и, помимо этого, уважению основного человеческого достоинства. Рассматриваемые права были выражены как свобода познания, неприкосновенность психического здоровья, психическая целостность и психологическая непрерывность.

8. Поэтому Ассамблея призывает государства-члены Совета Европы:

8.1. Установить этические рамки для исследований, разработки и применения нейротехнологий, включая технологию ВСІ, с учётом принципов, изложенных в параграфе 6 настоящей резолюции;

8.2. Чётко определить пределы исследований, разработки и применения нейротехнологий, включая технологию ВСІ, с помощью конкретных правовых рамок, обеспечивающих эффективное соблюдение и защиту прав человека;

8.3. Обеспечить наличие соответствующих органов для надзора и регулирования исследований, разработок и применения нейротехнологий, включая технологию ВСІ с тем, чтобы гарантировать эффективное внедрение применимых этических и правовых рамок;

8.4. Рассматривают создание и правовую защиту новых «нейрорайтов» как особенно эффективную защиту от возможных рисков, связанных с технологией ВСІ.

9. Что касается соответствующей работы, которая уже ведётся в рамках Совета Европы, Ассамблея:

9.1. Призывает ДН-ВІО к открытому и конструктивному подходу к вопросу о новых «нейроправах», включая возможность обеспечения их защиты в соответствии с международным правом посредством дополнительного протокола к Конвенции о защите прав человека и основных свобод (ETS № 005);

9.2. Призывает САНАІ учитывать потенциальные риски и возможности, возникающие в результате применения искусственного интеллекта в контексте систем ВСІ, и его особенно серьёзного воздействия на права человека.

## **В. Проект рекомендации**

1. Ассамблея ссылается на свою Резолюцию... (2020) под названием «Интерфейс мозг-компьютер: новые права или новые угрозы основным свободам?» Он напоминает, что эта резолюция была принята, поскольку соответствующая работа в рамках Совета Европы продолжалась Комитетом по биоэтике (ДН-ВІО), касающимся нейротехнологий, и Специальным комитетом по искусственному интеллекту (САНАІ).

2. Поэтому Ассамблея призывает Комитет министров:

2.1. Поддерживать работу в рамках ДН-ВЮ по правам человека и нейротехнологиям, в том числе путём дополнения существующего круга ведения, чтобы обеспечить рассмотрение возможности защиты «нейронных прав» посредством дополнительного протокола к Конвенции о защите прав человека и основных свобод (ETS № 005);

2.2. Принимать во внимание потенциально уникальное и беспрецедентное влияние на права человека использования искусственного интеллекта в связи с системами интерфейса мозг-компьютер (BCI) при оценке осуществимости правовой базы для искусственного интеллекта.

## **С. Пояснительный меморандум г-на Бехта, докладчика**

### **Введение**

1. Предложение о принятии резолюции (Дос 14814), лежащее в основе этого отчёта, которое я представил 23 января 2019 г., было передано в комитет 12 апреля 2019 г., после чего Комитет назначил меня докладчиком 29 мая 2019 г.

2. Это предложение напомнило о том, что компьютеры всегда считались олицетворением человеческого прогресса, поэтому постепенное соединение с человеческим телом устройств, использующих цифровые вычисления, происходило без особых дискуссий. Такие устройства теперь обеспечивают прямой интерфейс между человеческим мозгом и компьютерами. С одной стороны, это может иметь важные медицинские применения, такие как восстановление способности людей с ограниченными возможностями говорить или манипулировать физическими объектами; с другой стороны, он может позволить считывать, добавлять или удалять информацию из человеческого мозга, преодолевая барьер на пути к высшему прибежищу человеческой свободы: разуму.

3. После обсуждения Комитетом моего вступительного меморандума на заседании в Берлине 14-15 ноября 2019 года я совершил ознакомительный визит в Калифорнию, Соединенные Штаты Америки, 25-27 февраля 2020 года. В Калифорнийском университете в Беркли, я встретил профессора Джека Галланта (с которым я обсуждал последние разработки в области технологии визуализации мозга, включая ее использование для восстановления визуальных образов по нейронной активности), профессора Эхуда Исакоффа и профессора Стюарта Рассела; в Калифорнийском университете в Сан-Франциско я познакомился с Джошем Чартье, с которым обсудил использование электрокортикографии и алгоритмов машинного обучения «нейронной сети» для преобразования

декодированной нейронной активности в синтетическую речь. Оуэн Филлипс, генеральный директор начинающей компании BrainKey, рассказал мне о своих планах по разработке медицинских диагностических инструментов на основе искусственного интеллекта (ИИ) для анализа изображений магнитно-резонансной томографии (МРТ) платёжеспособных клиентов. Хосе Кармена, генеральный директор новой компании Iota Biosciences, рассказал мне о миниатюрных медицинских диагностических устройствах, которые иногда называют «нейронной пылью». В Стэнфордском университете я разговаривал с профессором Хэнком Грили, Аликс Роджер и Дэниелом Паланкером; Профессора Усама Хатиба и Шамика Гангули, с которыми я обсуждал робототехнику и протезирование; Профессор EJ Chichilnisky, с которым я обсуждал искусственные имплантаты сетчатки и технологические угрозы; и профессор Ник Мелош, с которым я обсуждал слуховые протезы. Профессор Байрон Ю из Университета Карнеги-Меллона описал последние разработки в области нейронной записи и технологии стимуляции, включая оптическую визуализацию. Алан Мардинли и Алекс Ферст из Neuralink рассказали о своей технологии нейронной записи. С Люком Джулией, вице-президентом по инновациям Samsung, я обсуждал потребительские приложения для взаимодействия с технологиями. С Грегом Коррадо, старшим научным директором Google, я обсуждал использование и возможные злоупотребления технологиями в целом и в сфере здравоохранения в частности. Я хотел бы поблагодарить всех этих людей за их время и вклад, а также поблагодарить Эммануэля Лебрена-Дамьена, генерального консула Франции в Сан-Франциско, и его коллег за их неоценимую помощь в организации моего визита.

4. Я также намеревался организовать слушание с экспертами на заседании Комитета, но, к сожалению, это стало невозможным из-за пандемии Covid-19. Я хотел бы поблагодарить д-ра Марчелло Иенка, заведующего кафедрой биоэтики, этики здоровья и лаборатории политики, D-HEST, Швейцарский федеральный технологический институт, ETH-Zurich, д-ра Тимоти Констандину, заместителя директора Центра биотехнологий, Имперский колледж Лондона и д-р Дэвид Виникофф, старший политический аналитик Организации экономического сотрудничества и развития (ОЭСР), за их готовность принять участие в запланированном слушании и за письменные вклады в этот отчет, которые они предоставили вместо этого.

## **Технологии**

5. История нейротехнологии относительно недавняя. Прошло чуть более ста лет с момента первой электроэнцефалографии (ЭЭГ), регистрирующей электрические сигналы в головном мозге животного; первая человеческая ЭЭГ была

зарегистрирована в 1924 году. Первая прямая электрическая стимуляция слуховой системы человека была проведена в 1957 году. В 1965 году ЭЭГ использовалась для сочинения музыки; к 1988 г. ЭЭГ можно было использовать для управления мобильным роботом. В 1997 году Управление по санитарному надзору за качеством пищевых продуктов и медикаментов США одобрило использование глубокой стимуляции мозга (DBS, инвазивная технология – см. Ниже) в качестве лечения эссенциального тремора и болезни Паркинсона. Прогресс ускорился за последние 20 лет, параллельно с экспоненциальным ростом вычислительной мощности: в 2005 году Мэтт Наги стал первым человеком, который управлял искусственной рукой с помощью интерфейса мозг-компьютер (BCI); в 2013 году пациент BrainGate управлял протезом конечности робота с помощью набора микроэлектродов, имплантированных в мозг; а в 2018 году исследователи из Беркли создали самый маленький и самый эффективный в мире имплантированный беспроводной нервный стимулятор «нейронная пыль». Ведущие игроки технологической индустрии сейчас работают над коммерческими приложениями, в том числе компания Neuralink Илона Маска и Reality Labs в Facebook.

6. Основная технология BCI состоит из двух компонентов: устройства для записи или стимуляции мозговой активности; и алгоритм «декодера» для извлечения информации из записанной активности или для создания сигнала, стимулирующего активность. Технологии записи и стимуляции могут быть либо неинвазивными (оставаясь вне черепа), либо инвазивными (внедряемыми в череп). Наряду с ЭЭГ, неинвазивные технологии регистрации включают функциональную магнитно-резонансную томографию (фМРТ) и функциональную ближнюю инфракрасную спектроскопию (fNIRS), которые регистрируют нервную активность путём измерения кровотока в головном мозге; Неинвазивные технологии стимуляции включают транскраниальную стимуляцию постоянным током (tDCS) и транскраниальную магнитную стимуляцию (TMS), которые индуцируют электрический ток в головном мозге. Технологии регистрации инвазивных ИМК включают электрокортикографию (ЭКоГ) и кортикальные имплантаты, которые включают размещение электродов непосредственно на коре головного мозга; на более экспериментальном уровне, «нейронная пыль» (беспроводные миниатюрные имплантаты без батарей, оснащённые датчиками и стимуляторами и питаемые от ультразвука), «нейронное кружево» (крошечные электроды, распределенные вдоль полимерных «нитей», которые вставляются – «вводятся») – в мозг, как это разрабатывается Neuralink) и «нейропиксели» (ещё один тип многоэлектродной матрицы, способной одновременно получать доступ ко многим различным областям мозга) представляют собой следующий шаг в усовершенствовании. Стимуляция глубокого мозга (DBS) и стимуляция блуждающего нерва (VNS) являются примерами технологий инвазивной стимуляции.

7. Во всех этих областях продолжается впечатляющий прогресс. Например, незадолго до того, как я приехал в Калифорнию, в Институте нейробиологии Хелен

Уиллс в Калифорнийском университете в Беркли, 7Т, был установлен сканер мозга NexGen 7T MRI, в связи с силой магнитов (7 тесла). МРТ NexGen 7Т является чрезвычайно дорогостоящим, как сообщается, 13,4 миллиона долларов) и действительно международным (компания Siemens, штаб-квартира которой находится в Германии, создала МРТ NexGen 7Т), как и многие другие исследования в этой области. МРТ NexGen 7Т будет иметь разрешение изображения около 0,4 мм, что соответствует масштабу структур нервных столбов, которые реагируют на определённые особенности сенсорного мира. В то время как пространственное разрешение, таким образом, становится все меньше ограничений, технология МРТ по-прежнему подвержена значительным ограничениям временного разрешения: в то время как нейроны срабатывают со скоростью около 100 раз в секунду, МРТ записывает со скоростью примерно один раз в секунду.

8. Несмотря на быстрый прогресс в последнее время и разнообразие подходов, которые в настоящее время исследуются, все ещё существует множество существенных ограничений для разработки более амбициозных приложений ВСИ. Неинвазивные подходы не могут регистрировать детальную активность на клеточном уровне и могут использоваться только для простых интерфейсов с двоичным выбором. В настоящее время все более сложное требует серьёзного хирургического вмешательства. Однако даже эти инвазивные технологии имеют серьёзные ограничения, включая ухудшение со временем качества записей, полученных с помощью имплантатов. Затем необходимо передавать данные изнутри черепа: либо через череп должен проходить провод, либо использовать беспроводную систему, которая создаёт свои собственные проблемы. Лучшие современные инвазивные системы с множеством электродов могут записывать до тысячи каналов, отслеживая сотни нейронов из одной области мозга; но ВСИ более общего назначения потребует отбор проб из десятков, если не сотен тысяч участков, возможно, из нескольких областей мозга. Это создаёт дополнительные проблемы для вычислений и анализа данных. Также существуют инженерные и хирургические проблемы, связанные с изготовлением и имплантацией сложных трёхмерных конструкций со встроенной электроникой. Прежде всего, «мы просто недостаточно хорошо понимаем природу распределенных представлений информации и обработки в неокортексе, чтобы иметь возможность сделать нечто большее, чем элементарную оценку того, что может «означать» конкретная последовательность действий».

9. По крайней мере, на технологическом фронте некоторые исследователи ожидают значительного прогресса в ближайшие годы на основе того, что они называют «нейронной нанороботикой». Это будет включать микроскопические устройства, которые будут вводиться в кровоток субъекта, за которым они будут следовать, при необходимости, пересекая гематоэнцефалический барьер, прежде чем обнаруживать и прикреплять к конкретным типам клеток мозга или клеточной структуре. Затем эти устройства будут записывать и / или стимулировать

нейронную активность, чтобы «обеспечить неразрушающую, работающую в реальном времени, безопасную, долгосрочную и практически автономную систему *in vivo*, которая может реализовать первый функциональный ИМК» – другими словами, система, которая преодолевает ограничения, упомянутые выше. «Такие человеческие ВСИ-системы могут кардинально изменить взаимодействие человека и машины, обещая значительное улучшение когнитивных функций человека».

10. Эти предложения не так надуманы, как может показаться. Электромагнитные наночастицы уже использовались для управления собственными полями в мозгу мыши, а флуоресцирующие углеродные наноточки использовались для нацеливания и визуализации определённых клеток в мозге мыши. Человеческий мозг был связан со спинным мозгом анестезированной крысы, а человеческий мозг управлял движениями таракана по S-образной дорожке с помощью электрической стимуляции его антенн. Множественные мозги были связаны для выполнения совместных задач: было обнаружено, что четыре соединённых мозга крысы превосходят один мозг по производительности вычислений; а интерфейс «мозг-мозг для трех человек» (BBI), называемый «BrainNet», позволял трём людям взаимодействовать с помощью неинвазивной прямой связи между мозгом (ЭЭГ и TMS) для принятия решений. Как говорится в документе Королевского общества о предполагаемых нейротехнологиях будущего: «До этих вещей еще далеко, но они не невозможны в какой-то форме. Подумайте, какой далёкой и футуристической казалась бы высадка на Луну или Интернет в 1950 году, когда у немногих семей даже был телефон...»

11. Прогресс в технологии ВСИ также был в значительной степени обусловлен быстрым развитием искусственного интеллекта за последнее десятилетие или около того. Анализ изображений мозга, особенно тех, которые получены с помощью МРТ, с помощью алгоритмов машинного обучения способствовал нашему пониманию того, как мозг функционально структурирован, и нашей способности декодировать нейронную активность, чтобы реконструировать образ мыслей, который он представляет. Это позволило исследователям реконструировать изображения из снимков МРТ субъектов, просматривающих трейлеры к фильмам (как это сделали профессор Галлант и его команда ещё в 2011 году), или преобразовать сигналы мозга, полученные с помощью ЭКГ, в синтетическую речь (как это делает доктор Чартье – см. Далее ниже). Как и в других областях, таких как медицинская диагностика или автономные транспортные средства, прогресс в области ИИ сыграл важную роль. Кроме того, общая информация об ИИ, включая описание и анализ применимых этических принципов, может быть найдена в приложении к настоящему отчёту.

## Приложения

12. Как отмечалось выше, самые ранние приложения ВСИ использовали сигналы мозга для управления простыми протезами и роботизированными устройствами. Примечание. По мере развития сенсорных устройств и вычислительной мощности стали возможны новые, более сложные приложения.

13. С учётом того, что исследования психологических расстройств и гериатрических неврологических расстройств, например, привлекают огромное количество средств, неудивительно, что многие современные исследования в области нейротехнологии сосредоточены на медицинских приложениях. Джош Чартье из Калифорнийского университета в Сан-Франциско рассказал мне о проекте, над которым он работал, по обнаружению паттернов нейронной активности, связанной с контролем голосового тракта во время речи. Затем полученные сигналы можно декодировать и использовать для генерации «синтетической речи». Такой подход позволяет обойти огромную трудность определения нейронной активности, связанной с конкретными словами как таковыми (по сравнению с активностью, связанной с намерением произнести эти слова). Исследование проводится на пациентах с тяжёлой формой эпилепсии, которым была установлена «нейронная сеть», предназначенная для отслеживания припадков, и которые выжили благодаря использованию устройства также для этого исследования. Такая нейротехнология синтеза речи может использоваться только людьми в самых экстремальных ситуациях, таких как «синдром запертости», поскольку она очень навязчива. Д-р Шартье признал возможность того, что техника, которая работает с людьми, которые все ещё могут говорить, может не работать для людей, которые долгое время не могли говорить и которые больше не могут демонстрировать те же подробные, последовательные паттерны нейронной активности; хотя он предположил, что с практикой пациенты с синдромом «запертости» могут «переобучить» свой мозг для создания моделей активности, которые можно расшифровать и использовать для генерации «синтетической речи». Основная концепция – получение артикулированного языка на основе нейронной активности – показывает, как растёт понимание того, как работает мозг, в сочетании с улучшением технологии чтения мозга. Интрузивные технологии и использование сознательных умственных усилий для управления голосовым трактом могут означать, что некоторые из этических проблем с точки зрения конфиденциальности и психической целостности, связанные с этим подходом, менее распространены.

14. Компания Хосе Кармены, Iota Biosciences, разрабатывает безбатарейные биоэлектронные устройства с ультразвуковым питанием, известные как «нейронная пыль», длиной всего несколько миллиметров, которые могут напрямую взаимодействовать с центральной нервной системой. Эти устройства могут собирать

точные данные или непосредственно стимулировать нервы и могут использоваться для диагностики и лечения состояний от артрита до сердечно-сосудистых заболеваний. Iota стремится уменьшить размер своих устройств до субмиллиметрового уровня, что в конечном итоге может сделать возможным их использование в технологии интерфейса мозг-компьютер.

15. Некоторые компании уже предоставляют услуги или продукты напрямую потребителю. BrainKey, который был представлен мне генеральным директором компании Оуэном Филлипсом, собирает базу данных снимков МРТ, в основном полученных из общедоступных баз данных (таких как UK Biobank), а также других данных от частных клиентов. В настоящее время услуга Брейки носит описательный / аналитический характер, включая статистику и трехмерную модель мозга человека. Намерение состоит в том, чтобы разработать инструменты диагностики AI. Хотя доктор Филлипс был против того, чтобы работодатели могли требовать от соискателей работы прохождения МРТ-сканирований, которые затем будут анализироваться в попытке предсказать будущую профессиональную деятельность человека, он согласился, что технология потенциально может быть использована для этой цели.

16. Канадская компания InteraXon Inc. продает мультисенсорное устройство на основе ЭЭГ под названием Muse, предназначенное для помощи пользователям во время практики медитации, регистрируя их нейронную активность (а также их физическое спокойствие с помощью акселерометра и их частоту сердечных сокращений с помощью оптического монитора сердечного ритма). Сопутствующее приложение для смартфона даёт пользователям отзывы об их практике и прогрессе. Пользователям сообщают, что они сохраняют полный контроль над данными ЭЭГ, сгенерированными при использовании их устройства Muse, но могут принять участие в исследовательской программе, которая передает анонимные данные ЭЭГ (и других датчиков Muse) «третьим лицам, участвующим в исследованиях, связанных с улучшением научных исследований. понимание мозга / тела или улучшение продуктов и / или предоставление лучшего опыта и услуг».

17. OpenBCI использует другой подход, продавая оборудование с открытым исходным кодом (включая ЭЭГ, наряду с ЭМГ для измерения мышечной активности и ЭКГ для функции сердца, а также всевозможные связанные компоненты), которое можно использовать с бесплатным программным обеспечением с открытым исходным кодом для различных проектов. На веб-сайте OpenBCI говорится, что «Мы работаем, чтобы использовать силу движения за открытый исходный код для ускорения этических инноваций в технологиях взаимодействия человека и компьютера». *Примечание. Это можно рассматривать как соответствующее модели «демократизации» нейротехнологического развития, предложенной д-ром Йенкой (см. далее ниже).*

18. Исследование, финансируемое Facebook, использовало технологию ЭКоГ, чтобы понять, на основе одних только нейронных сигналов, что испытуемые слышали и собирались сказать. Это было описано как «иллюстрация перспективных нейропротезных речевых систем для людей, которые не могут общаться»; обратите внимание, или, как заявил в Twitter вице-президент Facebook Эндрю Босуорт, «носимое устройство, которое позволяет людям печатать, просто представляя, что они хотят say «, что, возможно, более полно свидетельствует о коммерческом интересе Facebook к этой технологии. В то время как исследование доктора Шартье, например, уже дало многообещающие результаты с использованием инвазивной технологии ЭКоГ, трудно понять, насколько неинвазивные «носимые» устройства могут быть сравнительно эффективными, по крайней мере, в обозримом будущем.

19. Другой областью коммерческого интереса является область так называемого «нейромаркетинга», «недавняя междисциплинарная область, которая пересекает традиционные границы между нейробиологией, нейроэкономикой и маркетинговыми исследованиями... в первую очередь касающимися улучшения маркетинговых стратегий и продвижения продаж». Исследование Note One показало, что субъекты, которые предпочитали пепси-колу кока-коле во время слепой дегустации, имели сильную реакцию в одной области мозга (брюшная скорлупа); во время неслепой дегустации почти все испытуемые предпочитали кока-колу с особенно сильной реакцией в другой части мозга (префронтальной коре, которая связана с чувством себя и которая в данном контексте, казалось, подавляла вкусовые рецепторы тех, кто ранее предпочитал Pepsi). Более глубокое понимание этого механизма могло бы позволить рекламодателям разрабатывать и тестировать маркетинговые стратегии, предназначенные для задействования подсознательных механизмов предпочтений, которые особенно чувствительны к имиджу или бренду, а не к внутренним качествам продукта.

20. Нейротехнологии и ВСИ могут иметь особый потенциал в области уголовного и судебного разбирательства. «[Т] здесь попытки использовать нейробиологию для разработки объективных методов оценки того, что по своей сути является субъективным вопросом. Правдивая информация и обнаружение лжи, умственные способности, боль и надёжность памяти – полезные области исследования для уголовных и гражданских судебных процессов. Эти попытки чтения мыслей становятся теоретически возможными «. Примечание. Действительно, некоторые из этих технологий уже коммерчески доступны, даже если они ещё не приняты для использования в зале суда: например, компании No Lie MRI и Serphos продают свои продукты, использование фМРТ в качестве основы для оценки правдивости личности. Другое исследование показало, что активность в передней поясной части коры головного мозга, которая связана с импульсным контролем, была предиктором последующего повторного ареста (рецидивизма) среди правонарушителей. Примечание. Все эти события могут вызвать серьёзные вопросы с точки зрения основных прав и процессуальные гарантии.

21. Исследования нейротехнологии ВСИ проводятся не только в коммерческих и гражданских целях; это также делается для обеспечения безопасности и военных целей. Агентство перспективных исследовательских проектов Министерства обороны США (DARPA) проявляет особую активность в этой области. Большая часть его работы предназначалась в первую очередь для общих медицинских целей, хотя и с упором на вопросы, имеющие особое значение для вооружённых сил. Например, программы «революционного протезирования» и «технологии надёжных нейронных интерфейсов» (RE-NET) были предназначены для ускорения разработки протезов, контролируемых ИМК (нейропротезирование); и программа «реорганизация и пластичность для ускорения восстановления после травм» (REPAIR) была нацелена на восстановление нервных и поведенческих функций после нервной травмы или сенсорной депривации.

22. Другие исследования, финансируемые DARPA, могут вызвать более сложные этические вопросы. Программа «Нейронное устройство интеграции памяти с восстановительным кодированием» (REMIND) разработала технологию, которая может улучшить или ограничить способность субъекта записывать события в свою память. Программа «ускоренного обучения» была призвана «произвести революцию в обучении в военной среде» (например, в области стрельбы из винтовки). Программа «повествовательных сетей» (N2) была нацелена на обнаружение активности мозга, связанной с повествовательным влиянием, что могло бы позволить «более быструю и лучшую передачу информации в зарубежных информационных операциях» и создать технологии ВСИ, которые «замыкают петлю между сценаристом и потребителем, позволяя нейронные реакции на повествовательный стимул, чтобы диктовать траекторию рассказа». Это может быть использовано для создания «оптимальных повествований, адаптированных к конкретному человеку или группе людей». Программа «Нейротехнология для аналитиков разведки» (NIA) была предназначена для разработки новых неинвазивных систем ВСИ для повышения эффективности и продуктивности аналитиков изображений за счёт обнаружения нейронных реакций на видение «целей, представляющих интерес» на изображениях. «Система предупреждения об угрозах с помощью когнитивных технологий» (CT2WS) будет использовать аналогичный подход, чтобы повысить способность обнаруживать угрозы и реагировать на них во время наблюдения за безопасностью объекта. DARPA также работает над недорогими наушниками для ЭЭГ с целью «краудсорсинга усилий по сбору данных для нейровизуализации». Совсем недавно в рамках программы «Системная нейротехнология для новых методов лечения» (СУБСЕТИ) надеялись создать новые имплантируемые технологии для нейронной записи и стимуляции для лечения психоневрологических и неврологических состояний, в том числе среди ветеранов вооружённых сил с проблемами психического здоровья, связанными с службой. Программа «восстановления активной памяти» (RAM) направлена на восстановление памяти у людей с неизлечимыми заболеваниями, страдающих

дефицитом памяти. Возможно, из-за признания потенциальных этических проблем, DARPA, как сообщается, тесно сотрудничало с Управлением по санитарному надзору за качеством пищевых продуктов и медикаментов над этими двумя последними программами. Эксперты, с которыми я встречался в Калифорнии, считали, что проекты, финансируемые DARPA, далеки от достижения заявленные ими цели, которые часто намеренно выражались в чрезвычайно амбициозных терминах, с осознанием того, что они выходят за рамки того, что было возможно в настоящее время, и с ожиданием, что, тем не менее, будет достигнут хотя бы некоторый полезный прогресс.

23. Илон Маск сказал, что Neuralink призван найти способ «достичь своего рода симбиоза с искусственным интеллектом» – отражение его заявленного убеждения в том, что искусственный интеллект является «нашей самой большой угрозой существованию», «потенциально более опасным, чем ядерное оружие» (ядерное оружие). Эти опасения были поддержаны сторонниками «нейронанороботики» как средства взаимодействия человеческого мозга с облаком: это «может быть полезным для человечества, помогая смягчить серьёзные риски для существования, связанные с появлением общего искусственного интеллекта. [Это сделало бы это], позволив создать компенсирующую полезную технологию человеческого увеличения...»

24. Алан Мардинли, директор по исследованиям Neuralink, сказал мне, что видение Илона Маска «симбиоза» было далёкой от цели, и нужно было решить множество технических и научных неизвестностей, прежде чем её можно будет достичь. Непосредственной целью компании было производство медицинских устройств, в частности, для помощи пациентам с травмами спинного мозга. Поскольку внутрикортикальная технология «нейронного шнурка» Neuralink включала тысячи электродов, расположенных вдоль «нитей», свисающих с чипа 4x4 мм, она обеспечивала достаточное разрешение для управления устройствами. Г-н Мардинли считал нейронное кружево наилучшей доступной в настоящее время технологией для интерфейса мозг-компьютер, хотя его инвазивный характер означал, что спектр приложений обязательно будет узким. Тем не менее, он считал, что это предпочтительнее существующей стандартной электродной решётки штата Юта: нейронное кружево использовало гибкие электроды (в отличие от жёстких электродов массива Юты) и, таким образом, было менее разрушительным для ткани мозга и более долговечным, поскольку оно действительно не вызывает локальных рубцов; и он предлагал более высокое разрешение, от 8 до 16 электродов на очень узкую стойку (в отличие от одиночного электрода матрицы Юты на относительно толстой стойке). Со временем нейронное кружево можно будет использовать для имплантации до 100 000 электродов, хотя для этого потребуется интегрированная автоматизированная система введения. Подводя итог, Neuralink стремилась создать технологию, которая была бы безопасной, стабильной и масштабируемой. Neuralink уже работал с Управлением по контролю за продуктами и лекарствами США (FDA)

над «ранним технико-экономическим обоснованием»; в первую очередь это касалось вопросов безопасности, таких как кибербезопасность / возможность взлома, последствия для данных / конфиденциальности и проблемы контроля ввода, однако, а не этики.

25. Заглядывая дальше в (возможное) будущее, сторонники «нейронанороботики» предполагают, что эта технология, в сочетании с глобальной «облачной» инфраструктурой, будет использоваться во многих приложениях: полностью иммерсивная виртуальная реальность, которая будет неотличима от реальности; дополненная реальность, с наложением информации о реальном мире прямо на сетчатку; слуховой перевод на иностранные языки в реальном времени или доступ ко многим формам онлайн-информации; и даже испытать «полностью иммерсивные эпизоды в реальном времени из жизни любого желающего человека-участника на планете с помощью ненавязчивой «прозрачной тени».

26. В своем письменном вкладе в этот отчет доктор Винников изложил нижеприведенную таксономию нейротехнологий (NB, я немного изменил его предложение для простоты и актуальности). Хотя «интерфейс мозг-компьютер» проявляется в связи с основной функцией конкретной технологии, он, очевидно, также включает функции чтения и вмешательства / модуляции. Как уже отмечалось в этом отчете, технология ВСІ демонстрирует потенциал для использования во всех предлагаемых сферах.

Сферы использования, в которые входят:

- – клиника / медицина (неврология / нейрохирургия, психиатрия, реабилитация, медицина боли)
- – род занятий (обучение, выступление)
- – военные (разведка / допрос, улучшение оружия / истребителя)
- – общественные (непосредственно потребителю /сделай сам: образование, благополучие / образ жизни / развлечения)

Основная функция технологии, в том числе:

- – чтение мозга (визуализация, моделирование / картирование)
- – вмешательство / модуляция функции мозга
- – «Инженерия» мозга (интерфейс мозг-компьютер, нейропротезирование) производная (вдохновлённая мозгом, например, искусственные нейронные сети) Цели здоровья – предотвращать, восстанавливать, заменять, увеличивать...

27. Как резюмируется документ Королевского общества:

«Имплантаты, шлемы, повязки на голову или другие устройства могут помочь нам больше запоминать, быстрее учиться, быстрее принимать более правильные решения и решать проблемы без предубеждений... Связывание человеческого мозга

с компьютерами с использованием возможностей искусственного интеллекта может позволить людям объединить решения. создание эмоционального интеллекта людей с помощью вычислительной мощности компьютеров, создание новой формы коллективного интеллекта. Люди могут в какой-то степени стать телепатами, способными говорить не только без слов, но и без слов – через доступ к мыслям друг друга на концептуальном уровне. Не только мысли, но и чувственные переживания могут передаваться от мозга к мозгу... Умственно и физически улучшаться военнотружущие или полицейские могли бы защитить общественность, имея возможность более эффективно видеть в темноте, ощущать присутствие других и быстро реагировать...»

28. Тем не менее, все же стоит иметь в виду, что современные технологии очень далеки от достижения почти любой из этих целей или, по крайней мере, не имеют практической пользы (или угрозы). Без значительного прогресса в нашем понимании структуры и функционирования мозга наиболее впечатляющие возможности ВСИ останутся чисто умозрительными. У некоторых экспертов есть фундаментальные сомнения относительно потенциала нейротехнологии как основы для взаимодействия с компьютерами. Грег Коррадо из Google утверждал, что нервная система человека эволюционировала, чтобы взаимодействовать с внешним миром с помощью естественных органов чувств, которые останутся основными средствами взаимодействия с компьютерами: в частности, глаза и уши для получения информации, а также кончики пальцев и голосовые тракты для передачи информации. Таким образом, усовершенствования в интерфейсе «человек-компьютер» будут происходить на компьютерной стороне, например, обработка естественной речи (способность компьютера «понимать» человеческую речь). По его мнению, наиболее многообещающими возможностями для биоэлектронного интерфейса была прямая стимуляция мышечных нервов, кохлеарных и ретинальных имплантатов для ввода информации (стимуляции) и двигательной коры для считывания информации. По мнению доктора Коррадо, нынешнее состояние технологии ВСИ не создаёт никаких этических рисков, и эту точку зрения разделяли многие, с кем я встречался в Калифорнии. Тем не менее, его неправильное использование могло вызвать этические проблемы, например, если бы технология фМРТ использовалась для предсказания личности, преступных наклонностей или профессиональных способностей, что доктор Коррадо считал сродни френологии, дискредитированной теории, согласно которой индивидуальные черты и интеллектуальные способности человека можно определить путём измерения физических характеристик черепа.

## Проблемы

29. Текущее состояние технологии ВСІ может не вызывать непосредственных опасений, которые явно выходят за рамки существующих этических рамок (в частности, правил медицинской этики и конфиденциальности / защиты данных), но темпы их развития будут выходить за рамки этических проблем, и критическая точка, за которой действующее законодательство становится неадекватным, близка. Марш технического прогресса неумолим, и некоторые участники неизбежно будут стремиться разрабатывать приложения, которые действительно поднимают серьёзные этические проблемы, будь то на основе новых технологий или за счёт использования существующих технологий. Это не что иное, как простая мера предосторожности, позволяющая предвидеть то, что может произойти в будущем, и устанавливать ограничения, которые уведут исследования от предсказуемо вредных или опасных областей и к положительным приложениям, которые не угрожают индивидуальным правам или демократическим обществам. Некоторые из этих этических вопросов выходят за рамки настоящего отчёта: например, тот факт, что взаимодействие с мозгом «неизбежно изменяет мозг». Примечание. Для настоящих целей я сосредоточусь на вопросах, связанных с правами человека, которые основаны на в основной заботе об обеспечении уважения человеческого достоинства и основных характеристик человека как автономного морального агента.

30. Доступ к нейронным процессам, лежащим в основе сознательного мышления, подразумевает доступ к уровню «я», который по определению не может быть сознательно скрыт или отфильтрован – абсолютное нарушение конфиденциальности. Даже сегодняшние нейромаркетинговые технологии в сочетании с широкомасштабным распространением целевых сообщений через рекламу и социальные сети могут иметь разрушительное влияние на свободу выбора и демократические процессы. ВСІ могут быть использованы для создания продвинутых детекторов лжи, которые могут считаться настолько надёжными, что получаемая ими информация будет использоваться в качестве доказательств в уголовном процессе. Однако это может нарушить защиту от самообвинения: в действительности подозреваемый не может отказаться предоставить информацию следователю. В то же время ВСІ могут использоваться для создания ложных воспоминаний, изменения или удаления реальных, делая человеческие свидетельства ненадёжными в качестве доказательства.

31. Исследования «двойного назначения», финансируемые DARPA (см. Выше), вызывают особые опасения. Двойное использование относится как к использованию гражданских технологий в военных целях, для целей национальной безопасности или полиции / правосудия, так и к возможности вредоносного неправомерного использования, в том числе негосударственными субъектами,

использования полезных в иных отношениях технологий. В этом отношении было отмечено, что «зависимость этих технологий от вычислений и обработки информации также делает их потенциально уязвимыми для кибератак». Обратите внимание, чьи потенциальные последствия для человека становятся все более опасными по мере того, как технология ВСІ становится более сложной и её способность влиять на нейронные процессы более мощно и специфично.

32. Есть также вопросы доступа и справедливости. Уже появляются риски того, что доступ к мощным средствам улучшения когнитивных функций в неклинических целях может зависеть от благосостояния, причем одни могут себе это позволить, а другие нет. Это могло создать две категории людей: улучшенные и неусиленные. Эти технологии также таковы, что решения о том, что возможно, а кто должен получать выгоду, могут быть оставлены на нерегулируемый рынок или продиктованы интересами потенциально автократических правительств.

33. Наконец, «на более философском уровне существуют опасения, что широкое использование нейронных интерфейсов может привести к тому, что человеческие решения будут направляться тем, что некоторые называют» нейро-эссенциализмом» – восприятием того, что мозг является определяющей сущностью человека и что наш выбор может быть сведён к набору нейробиологических процессов, не оставляя места для индивидуальной деятельности или моральной ответственности...»

*Примечание. Можно спросить, в чем смысл разработки ВСІ как способа защиты человечества от потенциального риска общего AI, если при этом мы отрицаем определяющие качества индивидуального человеческого существования?*

34. Профессор Рафаэль Юсте из Колумбийского университета в Нью-Йорке сформулировал эти опасения от имени «Морнингсайдской группы», состоящей из 25 экспертов, работающих в области нейротехнологии и вокруг неё. В своём выступлении на заседании Межпарламентского союза в Дохе (Катар) в 2019 году профессор Юсте описал, как возможное злоупотребление и отсутствие регулирования могут привести к проблемам в следующих пяти областях:

- Личная идентичность: «чем больше мы подключаемся к сети через мозговые компьютерные интерфейсы или устройства, тем больше мы размываем нашу собственную идентичность, себя. Зависимость, которую мы наблюдаем сейчас, от наших устройств, является аппетитом к тому, что должно произойти: по мере того, как мы увеличиваем пропускную способность нашего подключения к сети, используя неинвазивные мозговые компьютерные интерфейсы, мы все больше растворяемся в ней».
- Свобода воли: «если мы используем внешние алгоритмы и информацию для принятия решений, мы отказываемся от собственной свободы воли. Кто

принимает решение? ... Что произойдёт, когда у нас будет GPS-навигатор, который подсказывает нам, что мы должны делать в любой момент».

➤ Психическая конфиденциальность: «Если данные мозга доступны и могут быть расшифрованы, тогда наши умственные процессы, наши мысли будут доступны извне. Более того, можно расшифровать даже мысли, о которых мы не подозреваем или не подсознательно. Мы считаем, что данные о мозге должны быть защищены с той же законодательной строгостью, что и органы тела. Фактически, данные нашего мозга – это орган, не физический, а ментальный, и следует запретить торговлю с ними, поскольку они представляют, кто мы есть».

➤ Когнитивное расширение, включая расширенное обучение, которое «могло бы позволить некоторым группам общества в некоторых странах увеличить свои умственные и физические способности, дав им доступ к внешним алгоритмам и робототехнике для повседневной жизни. Мы считаем, что обеспечение принципа справедливости при разработке и внедрении этих технологий должно гарантировать равенство доступа и что использование этих технологий в военных целях должно строго регулироваться».

➤ Защита от предубеждений и дискриминации, «поскольку алгоритмы, используемые в ИИ, часто имеют неявные предубеждения, поэтому эти технологии могут непреднамеренно внедрить эти предубеждения в обработку нашего мозга. Будет ужасно разрушить наш исторический путь к равенству и справедливости, распространяя предубеждения с помощью новых технологий».

## 5 ответов

35. Среди исследователей широко распространено мнение о необходимости упреждающих регуляторных действий в отношении появляющихся нейротехнологий, включая ВСИ: то, что авторы статьи Королевского общества называют «ранним и часто» подходом. Промышленность также поддерживает такой подход: Марк Шевиче, директор по исследованиям в Facebook Reality Labs, заметил, что «мы не можем предвидеть или решить все этические проблемы, связанные с этой технологией, в одиночку. Что мы можем сделать, так это распознать, когда технология продвинулась дальше того, что люди знают о возможностях, и убедиться, что информация доставляется обратно сообществу. Нейроэтический дизайн является одним из ключевых столпов нашей программы – мы хотим быть прозрачными в том, над чем мы работаем, чтобы люди могли сообщить нам о своих опасениях по поводу этой технологии». Примечание. Другие выступали за координацию исследований в области нейроэтики и смежных областях, например, ИИ – совпадение двух областей с точки зрения технологий должно быть отражено в общем подходе к этическим принципам. Примечание. Некоторые высказывались в

пользу «структуры нейробезопасности...», разработанной и реализованной для обеспечения максимальной безопасности во всем трансляционном континууме между научными исследованиями и обществом (и наоборот)... особенно чувствительны к предвидению и своевременному обнаружению специфических для нейротехнологий угроз, особенно тех, которые касаются ментального измерения... [Эта структура] должна включать, по крайней мере, три основных уровня защиты: калиброванные регуляторные вмешательства, коды этического поведения и информационно-просветительской деятельности «.

36. В 2013 году Совет Наффилда по биоэтике опубликовал подробный отчет, озаглавленный «Новые нейротехнологии: вмешательство в мозг». В этом отчете, который фокусируется в первую очередь, но не исключительно на медицинских приложениях или нейротехнологиях, отмечается, что «мозг имеет особый статус в жизни человека, который отличает его от других органов. Его здоровое функционирование играет центральную роль в функционировании наших тел, наших способностях к автономной деятельности, наших представлениях о себе и наших отношениях с другими – и, следовательно, в наших способностях вести полноценную жизнь. Это означает, что новые нейротехнологии, которые мы рассматриваем в этом отчете, каждая из которых воздействует на мозг, вызывают этические и социальные проблемы, которые не возникают в такой же степени другими биомедицинскими технологиями». В отчете предлагается этическая основа, состоящая из трёх этапов:

- Основополагающие принципы благотворительности и осторожности, возникающие из «противоречия между потребностью и неопределённостью». В то время как серьёзные заболевания головного мозга вызывают серьёзные страдания, отсутствуют (другие) эффективные меры. В то же время все преимущества и риски нейротехнологий также ещё не полностью изучены из-за их новизны и неполного понимания того, как работает сам мозг. «Таким образом, особый статус мозга дает как повод проявлять благосклонность, вмешиваясь, когда травма или болезнь вызывают расстройства мозга, так и повод для осторожности, когда мы не уверены, каковы будут последствия этого».
- Значение принципов милосердия и осторожности следует рассматривать с учетом пяти интересов: безопасность, (непреднамеренное) воздействие на частную жизнь, автономность (как в принятии решений, касающихся лечения, так и в более широком контексте жизни пациентов), справедливость доступ к новым методам лечения и доверие к новым технологиям.
- В докладе также предлагаются три «достоинства», которыми руководствуются участники в этой области. Это изобретательность (как в инновациях, так и в обеспечении более широкого доступа), скромность (признание ограниченности знаний и нашей способности использовать технологии для облегчения заболеваний мозга) и ответственность (за счет

использования надежных исследовательских практик и воздержания от преувеличенных или преждевременных заявлений. для нейротехнологий).

37. Подобные подходы отражены в других предложениях по этическим рамкам для регулирования ВСИ и других нейротехнологий. Например, одна группа исследователей предложила следующий список «Руководящих принципов нейроэтики»:

I. Сделать оценку безопасности превыше всего;

II. Предвидеть особые проблемы, связанные с возможностями, автономией и агентством;

III. Защитить конфиденциальность и конфиденциальность нейронных данных;

IV. Заботиться о возможном злонамеренном использовании инструментов нейробиологии и нейротехнологий;

V. с осторожностью переносить инструменты нейробиологии и нейротехнологии в медицинские или немедицинские цели;

VI. Выявление и решение конкретных проблем общества в отношении мозга;

VII. Поощрять просвещение общественности и диалог;

VIII. Ведите себя справедливо и пользуйтесь преимуществами исследований в области нейробиологии и полученных в результате технологий.

38. Проблема «двойного использования когнитивных технологий» (включая ВСИ) была затронута Марсело Йенка в статье 2018 года. Отметив, что «когнитивные технологии обладают потенциалом для ускорения технологических инноваций и приносят значительную пользу отдельным людям и обществу», д-р Янка считает, что «благодаря их потенциалу двойного назначения, они могут использоваться государственными и негосударственными субъектами в неблагоприятных целях, включая киберпреступность, кибертерроризм, кибервойну и массовое наблюдение. В свете недавнего глобального кризиса демократии, усиления милитаризации цифровой инфосферы и одновременного усиления когнитивных технологий [КТ] важно проактивно разрабатывать стратегии, которые могут снизить возникающие риски и согласовать будущее КТ с основными принципами либеральной демократии в свободных и открытых обществах».

39. Поэтому д-р Ienca's предлагает «демократизацию СТ», приняв элементы двух крайностей – управления и регулирования, чистого невмешательства и строгого регулирования. «Демократизация» разделяет со строгим регуляторным подходом оценку новизны КТ как относительно недавней и все еще незрелой области, в которой отсутствует консенсус по основным концепциям или политике, которые потребуются для максимизации выгод при минимизации рисков. В нем

также признается масштаб возможностей и рисков – «потенциал влияния на когнитивные способности человека, определяющий немаловажное влияние на культурную эволюцию человека и глобальное равновесие» – и тот факт, что новизна КТ означает, что «человеческие общества сейчас находятся на историческом этапе, когда они могут принимать упреждающие решения о типе сосуществования, которое они хотят установить с этими технологиями. Привилегированные подходы к невмешательству на этом этапе развития отложили бы вмешательство по управлению рисками до того времени, когда когнитивные технологии широко развиты и широко используются, а значит, не поддаются модификации» Демократизация, однако, согласилась бы с точкой зрения невмешательства, что «чрезмерное регулирование может (а) свести на нет преимущества когнитивных технологий для общества в целом, и, если оно будет управляться недемократическими или несовершенными демократическими правительствами, (б) приведет к нежелательному концентрации власти и контроля».

40. Эта «демократизация» КТ будет основана на шести нормативных принципах, которые сопоставимы с компонентами этических рамок, предлагаемых для регулирования нейротехнологий в более общем плане:

I. Избегание централизованного контроля – «принцип, согласно которому с моральной точки зрения предпочтительнее избегать централизованного контроля над СТ для предотвращения рисков, связанных с неограниченным накоплением капитала, власти и контроля над технологией среди организованных групп, таких как крупные корпорации или правительства... Нормативные вмешательства, направленные на ограничение этого риска централизации, могут быть концептуализированы как киберэтические аналоги антимонопольных законов».

II. Открытость – «принцип содействия всеобщему доступу к (компонентам) дизайна или плана когнитивных технологий и всеобщему перераспределению этого дизайна или плана через открытый и совместный процесс совместного производства». Избегание централизованного контроля (Принцип I выше) и открытость «являются критически важными требованиями для того, чтобы те же самые возможности, которые будут записаны с помощью когнитивных технологий или внедряться в них доступны каждому». «В более абстрактном смысле открытость в компьютерной томографии подразумевает принцип внедрения в каждое приложение, с которым мы взаимодействуем, на любом устройстве в любой момент времени с (компонентами) когнитивных технологий».

III. Прозрачность – «принцип, позволяющий широкой общественности понять внутренние процессы когнитивных технологий».

IV. Инклюзивность – «принцип обеспечения того, чтобы никакая группа лиц или меньшинств не была изолирована или оставлена позади в процессе проникновения когнитивных технологий в наше общество... Принцип

инклюзивности [применяется к любому] этически значимому социальному предубеждению, которое может возникнуть намеренно или непреднамеренно во время разработки КТ. К ним относятся культурные, политические, языковые предубеждения и т. д.»

V. Ориентация на пользователя – «новые когнитивные технологии должны разрабатываться, разрабатываться и внедряться в соответствии с потребностями пользователей и личным выбором... конечных пользователей (как можно более широко охарактеризованных в соответствии с принципами открытости и инклюзивности) [должны быть] вовлечены в проектировании, разработке и внедрении когнитивных технологий на равных».

VI. Конвергенция. «В узком смысле, конвергенция – это принцип функциональной совместимости, взаимодействия и простоты интеграции между всеми компонентами когнитивных технологий ... [хотя] чрезмерная совместимость может привести к повышенной небезопасности данных ... В более широком и абстрактном смысле это также принцип конвергенции различных типов когнитивных технологий, особенно нейротехнологий, с одной стороны, и систем искусственного интеллекта, с другой». 41. Что касается нейротехнологий, разработанных специально для военных приложений, утверждалось, что «Хотя глобальный запрет или мораторий на военные нейротехнологии в настоящее время кажется этически неоправданным, могут потребоваться более мягкие и более взвешенные регулирующие меры для снижения рисков непропорционального использования нейробиологии в качестве оружия» Опять же, это будет означать «острую необходимость в мониторинге и тщательной оценке рисков в контексте технологий двойного назначения. Несмотря на то, что в настоящий момент кажется, что польза перевешивает риски, должны быть предусмотрены превентивные механизмы, позволяющие оперативно обнаруживать будущие изменения соотношения риска и пользы». Примечание. Тем не менее, необходимо не лишать себя возможности защитить себя в событие, когда злоумышленник использует эти технологии.

42. Между областью «нейроэтики» и областью биоэтики существует много общего. В области биоэтики в 1997 году Совет Европы принял Конвенцию о защите прав человека и достоинства человека в связи с применением достижений биологии и медицины: Конвенция о правах человека и биомедицине (СЕД № 164, «Конвенция Овьедо»), цель которой – «защитить достоинство и идентичность всех людей и гарантировать каждому, без дискриминации, уважение их неприкосновенности и других прав и основных свобод в отношении применения биологии и медицины». Среди прочего, Конвенция Овьедо гласит, что «интересы и благополучие человека превалируют над интересами общества или науки»; обязывает Стороны «принимать соответствующие меры с целью обеспечения в пределах их юрисдикции равноправного доступа к медицинской помощи надлежащего качества»; и

устанавливает, что «любое вмешательство в области здравоохранения, включая исследования, должно осуществляться в соответствии с соответствующими профессиональными обязательствами и стандартами». Он также содержит подробные положения о согласии, допустимом использовании прогностических тестов и допустимых целях вмешательства в геном человека.

43. В ноябре 2019 года, вскоре после того, как я представил свой вступительный меморандум комитету, межправительственный комитет Совета Европы по биоэтике ДН-ВЮ принял Стратегический план действий по правам человека и технологиям в биомедицине (2020-2025 годы). В этом документе отмечается, что «Применение новых и конвергентных технологий в биомедицине приводит к стиранию границ между физическими и биологическими науками, между лечением и исследованиями, а также между медицинскими и немедицинскими целями. Несмотря на то, что они предлагают значительные возможности в области биомедицины и за ее пределами, они также поднимают новые этические проблемы, связанные, в частности, с идентичностью, автономией, конфиденциальностью и недискриминацией». Основываясь на сотрудничестве (между органами Совета Европы и с другими соответствующими межправительственными органами) и коммуникации (с внешними заинтересованными сторонами), План действий структурирован вокруг трех столпов: управление (основанное на правах человека, общественный диалог, демократическое управление и прозрачность), справедливость (в доступе к инновационным методам лечения и технологиям для борьбы с неравенством в отношении здоровья из-за социальных и демографических изменений) и целостность (включая расширение участия детей, защиту прав детей и защиту прав лиц с проблемами психического здоровья).

44. В отношении ВСІ и связанных с ним технологий в Плане действий отмечается, что «разработки в области нейротехнологий, такие как глубокая стимуляция мозга, интерфейсы мозг-компьютер и искусственные нейронные сети, повышают перспективы более глубокого понимания, мониторинга, а также контроля над человеческим мозгом, поднимая вопросы неприкосновенности частной жизни, личности и дискриминации ... Следовательно, необходимо оценить, могут ли эти проблемы быть в достаточной степени решены с помощью существующей системы прав человека или новые права человека, относящиеся к когнитивной свободе, психической конфиденциальности, а также психической целостности и психологической преемственности, нужно развлекаться, чтобы управлять нейротехнологиями. В качестве альтернативы другие гибкие формы надлежащего управления могут лучше подходить для регулирования нейротехнологий». Неудивительно, что ДН-ВЮ разделяет мои опасения и общие черты моего восприятия возможностей и рисков, которые представляет технология ВСІ. Я приветствую и поддерживаю его будущую работу по ключевым ответным мерам, а именно целевому укреплению системы прав человека (см. ниже) и разработке

гибкого режима регулирования, который может поддерживать и направлять исследования и разработки для достижения позитивных, конструктивных целей.

45. Другие международные организации также внимательно относятся к новым нейротехнологиям. В декабре 2019 года Совет ОЭСР принял Рекомендацию об ответственных инновациях в нейротехнологиях. В этой рекомендации «подчеркивается важность (1) ценностей высокого уровня, таких как руководство, доверие, безопасность и конфиденциальность в этом технологическом контексте, (2) наращивания потенциала ключевых институтов, таких как органы прогнозирования, надзора и консультирования, и (3) процессы общественного обсуждения, инклюзивных инноваций и сотрудничества». Затем он призывает государства-члены и государства, не являющиеся членами, а также «всех участников» продвигать и внедрять серию из девяти «принципов ответственных инноваций в нейротехнологии», в которых подробно описаны конкретные действия. Принципы:

I. Содействие ответственным инновациям;

II. Приоритетность оценки безопасности;

III. Поощрение инклюзивности;

IV. Развитие научного сотрудничества;

V. Обеспечение общественного обсуждения;

VI. Расширение возможностей надзорных и консультативных органов;

VII. Защита личных данных мозга и другой информации;

VIII. Поощрение культуры управления и доверия в государственном и частном секторе;

IX. Предвидение и мониторинг потенциального непреднамеренного использования или неправильного использования.

46. Нейротехнологии в целом и ВСИ в частности могут фундаментально изменить отношения между внутренним «я» и внешним миром. Поэтому исследователи (в том числе д-р Йенка) призвали к новаторским правовым ответам, включая создание (или уточнение) четырех «новых» обеспеченных правовой санкцией прав человека: право на когнитивную свободу, право на неприкосновенность частной жизни, право на психическую неприкосновенность и право к психологической преемственности. «Когнитивная свобода» может рассматриваться как «право и свобода контролировать свое собственное сознание, а электрохимические мыслительные процессы [и как таковые] являются необходимой основой практически для любой другой свободы». В этом отношении его можно сравнить со свободой мысли, которую можно рассматривать как необходимое условие свободы религии, выражения мнений и ассоциаций. Сравнимо, но

фундаментально отличное и отличное: если свобода мысли – это право думать все, что угодно, то когнитивная свобода является ее предварительным условием – правом мозга генерировать мысли без технологического (или другого) вмешательства в этот процесс. Право на психическую неприкосновенность защитит людей от несогласованного наблюдения за их подсознательными психическими процессами. Право на психическую неприкосновенность защитит от вреда в форме «злонамеренного взлома мозга», давая контроль над мыслями и действиями человека. Право на психологическую преемственность защитит от действий, которые могут повлиять на «восприятие людьми своей собственной идентичности... [как] состоящих в переживании себя как сохраняющегося во времени одного и того же человека». Примечание – психологически оставаться самим собой. Статья, в которой представлены эти предложения, обширна, подробна и наводит на размышления. В нем рассматриваются вопросы, в том числе, подразумеваются ли эти «новые» права уже в существующих правах, и будет ли их «создание» равносильно «инфляции прав».

47. Одна страна, Чили, уже работает над правовой защитой «нейроправа» в сотрудничестве с профессором Юсте и инициативой NeuroRights Колумбийского университета (см. Выше). Предлагаемая поправка к статье 19 чилийской конституции будет определять психическую идентичность как основное право, которое может быть изменено только в соответствии с будущими законами. В сопутствующем законопроекте «Нейрозащита» будут установлены юридические определения нейротехнологии, интерфейсов мозг-компьютер и нейротехнологических органов. Все данные, полученные из мозга, будут определяться как нейроданные и подпадать под действие действующего законодательства о донорстве органов, тем самым запрещая торговлю нейроданными. Любое будущее использование и развитие нейротехнологии будет регулироваться медицинским законодательством. Наряду с этими инициативами Католический университет в Чили работает над этическими принципами для компьютерной индустрии, искусственного интеллекта и нейроинженерии. Все эти действия сочетаются с кампанией по информированию общественности, поддерживаемой президентом, министрами правительства и парламентариями. Примечание. Это сделает Чили первой страной в мире, которая будет регулировать и защищать данные, которые могут быть извлечены из человеческого мозга, чтобы эти данные могут использоваться только в альтруистических целях. Примечание. Утверждается, что новая чилийская правовая база сделает такие технологии, как проект Facebook по типу «от мысли к типу», незаконным.

48. Существует некоторый скептицизм по отношению к призывам нейроэтиков о правовой защите нейропсихиков. Алан Мардинли из Neuralink, например, выдвинул гипотезу, что право на познавательную свободу может быть нарушено обычной рекламой, которая часто преднамеренно создавалась и нацелена на использование подсознательных пристрастий; точно так же лечение от

зависимости можно рассматривать как внешнее вмешательство в свободу выбора человека в отношении потребления.

49. Что касается аспекта искусственного интеллекта в технологии ВСІ, в сентябре 2019 года Комитет министров учредил Специальный комитет по искусственному интеллекту (САНАІ). САНАІ было поручено изучить осуществимость и потенциальные элементы правовой базы для проектирования, разработки и применения искусственного интеллекта. Его работа основана на стандартах Совета Европы в области демократии, прав человека и верховенства закона, а также на других соответствующих международно-правовых документах и на текущей работе других международных и региональных организаций. Помимо обычных участников, представляющих государства-члены Совета Европы, государства-наблюдатели и другие органы Совета Европы (включая Ассамблею), САНАІ имеет исключительно высокий уровень участия представителей органов частного сектора, гражданского общества, исследовательских и академических институтов.

50. САНАІ провела своё первое заседание 18-20 ноября 2019 года. Среди прочего, он решил, что ключевым элементом будущего технико-экономического обоснования будет «отображение рисков и возможностей, возникающих в результате разработки, проектирования и применения искусственного интеллекта, включая влияние последнего на права человека, верховенство закона и демократию». В настоящее время САНАІ планирует принять технико-экономическое обоснование на своём третьем заседании, запланированном на декабрь 2020 года.

51. Это институциональный контекст, в котором Ассамблея будет обсуждать настоящий и различные другие отчёты, связанные с ИИ, которые в настоящее время готовятся в различных комитетах. Ассамблея решила подойти к этой теме на основе контекста, изучив влияние ИИ в различных областях. В рамках Комитета по правовым вопросам и правам человека, например, есть также отчёты о влиянии ИИ на «Правосудие по алгоритму – роль искусственного интеллекта в полицейской деятельности и системах уголовного правосудия», на «Правовые аспекты» автономных «транспортных средств. «И (на начальных этапах подготовки) летальных автономных систем вооружения. Таким образом, рекомендации, которые Ассамблея может принять на основе этих отчётов, послужат важным руководством для САНАІ при составлении карты рисков и возможностей ИИ и его воздействия на права человека, верховенство закона и демократию и последующем определении необходимости обязательного международно-правовая база.

## **Выводы и рекомендации**

52. Как и многие другие технологии, развитие ВСІ-технологии создаёт как возможности, так и риски. ВСІ могут использоваться для восстановления способности людей двигаться и общаться, сотрудничать при выполнении задач или выполнять их с большей эффективностью и эффективностью; для повышения когнитивных способностей за счёт прямого доступа к данным и использования дополнительных вычислительных мощностей; или испытать новые сенсорные или даже эмоциональные ситуации. С другой стороны, они могут использоваться для обхода прав на неприкосновенность частной жизни, неприкосновенности и защиты от самоговора, а также свободы выражения мнения; влиять на выбор, поведение и даже на долговременную личность; или подорвать фундаментальные характеристики человеческого равенства и достоинства.

53. Хотя ни научное понимание, ни технология не достаточно развиты, чтобы вызвать все эти опасности, некоторые из них уже возможны. Все приведённые выше примеры являются реалистичными, предсказуемыми последствиями быстрого прогресса. Как пришли к выводу многие комментаторы с различных точек зрения, существует неотложная и неотложная необходимость предвидеть потенциальные риски и принимать регулирующие меры для их смягчения или предотвращения. Как и в смежной области ИИ, это может принимать форму этических хартий, обязательных правил или даже новых прав; или, скорее всего, комбинация всех трёх. Технология ВСІ может все больше полагаться на ИИ, но сама по себе вызывает отдельный набор проблем. Поэтому требуемые этические принципы и ответные меры регулирующих органов в некотором роде более сложны, что отражает важность того, что означает вторжение этой технологии в самый центр нашего человеческого существа.

54. Мои практические и политические предложения изложены в прилагаемом проекте резолюции и рекомендации.

## **Предотвращение дискриминации, вызванной использованием искусственного интеллекта**

Отчёт | Док. 15151 | 29 сентября 2020 г.

Комитет Комитет по равенству и недискриминации

Докладчик: Г-н Кристоф ЛАКРУА, Бельгия, SOC

Происхождение Обращение в комитет: док. 14808, номер 4434 от 12 апреля 2019 г. 2020 г. - октябрь Постоянный комитет

## **Резюме**

Искусственный интеллект (ИИ), позволяя масштабировать автоматизированные процессы принятия решений, создаёт возможности для повышения эффективности – но параллельно он может закрепить и усугубить дискриминацию. Уже было показано, что использование ИИ в государственном и частном секторах оказывает дискриминационное воздействие, в то время как потоки информации, как правило, подчёркивают крайности и разжигают ненависть. Использование предвзятых наборов данных, дизайн, который не учитывает необходимость защиты прав человека, отсутствие прозрачности алгоритмов и ответственности за их влияние, а также отсутствие разнообразия в командах ИИ – все это способствует этому явлению.

Государства должны действовать прямо сейчас, чтобы ИИ не оказывал дискриминационного воздействия на наши общества, и должны работать вместе над разработкой международных стандартов в этой области.

Более того, парламенты должны играть активную роль в надзоре за использованием технологий на основе искусственного интеллекта и обеспечивать его общественный контроль. Внутреннее антидискриминационное законодательство должно быть пересмотрено и изменено, чтобы обеспечить жертвам дискриминации, вызванной использованием ИИ, доступ к эффективным средствам правовой защиты, а национальные органы по вопросам равенства должны быть эффективно оснащены для борьбы с воздействием технологий на основе ИИ.

Уважение к равенству и недискриминации должно быть интегрировано с самого начала в разработку систем на основе ИИ и проверено перед их развёртыванием. Государственный и частный секторы должны активно продвигать разнообразие и междисциплинарные подходы в исследованиях технологий и профессиях.

## **А. Проект резолюции**

1. Искусственный интеллект (ИИ) меняет наш образ жизни. Он позволяет масштабировать процессы и уже используется широким кругом частных и государственных организаций в самых разных областях, таких как процедуры отбора, определяющие доступ к занятости и образованию, оценка индивидуальных прав на социальные выплаты или кредиты или таргетирование рекламы и новостей.

2. Многие виды использования ИИ могут иметь прямое влияние на равенство доступа к основным правам, включая право на личную жизнь и защиту личных данных; доступ к правосудию и право на справедливое судебное разбирательство, в частности в отношении презумпции невиновности и бремени доказывания; доступ к занятости, образованию, жилью и здравоохранению; и доступ к общественным услугам и социальному обеспечению. Было обнаружено, что использование ИИ вызывает или усугубляет дискриминацию в этих областях, что приводит к отказу в доступе к правам, которые непропорционально затрагивают определённые группы - часто женщины, меньшинства и тех, кто уже является наиболее уязвимым и маргинализированным. Его использование в информационных потоках также связано с распространением онлайн-ненависти, которая распространяется на все другие социальные взаимодействия.

3. Машинное обучение, используемое для создания систем на основе искусственного интеллекта, основано на огромных наборах данных (больших данных), большая часть которых является персональными данными. В этом контексте необходимы эффективные гарантии защиты личных данных. В то же время данные по своей природе необъективны, поскольку они отражают дискриминацию, уже присутствующую в обществе, а также предвзятость тех, кто их собирает и анализирует. Выбор того, какие данные использовать, а какие игнорировать в системах на основе ИИ, а также отсутствие данных по ключевым вопросам, использование прокси и трудности, связанные с количественной оценкой абстрактных концепций, также могут привести к дискриминационным результатам. Предвзятые наборы данных лежат в основе многих случаев дискриминации, вызванной использованием ИИ, и остаются серьёзной проблемой, которую необходимо решить в этой области.

4. Дизайн и назначение систем на основе искусственного интеллекта также имеют решающее значение. Алгоритмы, оптимизированные для обеспечения эффективности, прибыльности или других целей, без должного учёта необходимости гарантировать равенство и недискриминацию, могут вызывать прямую или косвенную дискриминацию, включая дискриминацию по ассоциации, по широкому спектру признаков, включая пол, пол, возраст, национальное или этническое происхождение, цвет кожи, язык, религиозные убеждения, сексуальная ориентация, гендерная идентичность, половые характеристики, социальное происхождение, гражданское положение, инвалидность или состояние здоровья. Это делает особенно важным, везде, где их использование может повлиять на доступ к основным правам, чтобы системы на основе ИИ с самого начала включали в себя полное уважение равенства и недискриминации и подвергались тщательному тестированию, прежде чем они будут применены и развёрнуты, а также регулярно после их развёртывания эти права были гарантированы.

5. Сложность систем искусственного интеллекта и тот факт, что они часто разрабатываются частными компаниями и рассматриваются как их интеллектуальная собственность, могут привести к серьёзным проблемам с прозрачностью и подотчётностью в отношении решений, принимаемых с использованием этих систем. Это может чрезвычайно затруднить доказательство дискриминации и затруднить доступ к правосудию, в частности, когда бремя доказывания возлагается на потерпевшего и / или когда предполагается, что машина по умолчанию приняла правильное решение, что нарушает презумпцию невиновности.

6. Отсутствие разнообразия во многих технологических компаниях и профессиях увеличивает риск того, что системы ИИ будут разрабатываться без должного учёта их потенциально дискриминационного воздействия на некоторых людей и группы в обществе. Доступ женщин и меньшинств к профессиям в области науки, техники, инженерии и математики (STEM) необходимо улучшить, и в этой профессиональной среде срочно необходимо развить настоящую культуру уважения к разнообразию. Использование междисциплинарных и межкультурных подходов на всех этапах разработки систем ИИ также будет способствовать их укреплению с точки зрения равенства и недискриминации.

7. Наконец, строгие, ясные, общепринятые и применимые этические принципы должны лежать в основе разработки и развёртывания всех систем на основе ИИ. Парламентская ассамблея считает, что эти принципы можно сгруппировать под следующими широкими заголовками: прозрачность, включая доступность и объяснимость; справедливость и беспристрастность, включая недискриминацию; ответственность человека за принимаемые решения, включая ответственность и доступность средств правовой защиты, а также защиту и безопасность; и конфиденциальность, и защита личных данных.

8. Ассамблея приветствует тот факт, что как государственные, так и частные субъекты начали изучать и разрабатывать этические стандарты и стандарты прав человека, применимые к использованию ИИ. Он, в частности, приветствует Рекомендацию Комитета министров Res / CM (2020) 1 о влиянии алгоритмических систем на права человека, а также сопровождающие её руководящие принципы по устранению воздействия алгоритмических систем на права человека, а также рекомендацию Комиссара Совета Европы за права человека на тему «Распаковка искусственного интеллекта: 10 шагов для защиты прав человека». Он поддерживает общие предложения, содержащиеся в этих текстах, для применения также в области равенства и недискриминации.

9. Ассамблея подчёркивает, что законодатели не должны прятаться за сложностями ИИ, чтобы помешать им вводить правила, направленные на защиту и поощрение равенства и недискриминации в этой области: поставленные на карту вопросы прав человека ясны и требуют действий. Помимо этических принципов,

необходимы процедуры, инструменты и методы для регулирования и аудита систем на основе ИИ, чтобы обеспечить их соответствие международным стандартам в области прав человека, и в частности, правам на равенство и недискриминацию. Учитывая сильные транснациональные и международные аспекты технологий на основе ИИ, в этой области также необходимы международные стандарты.

10. В свете этих соображений Ассамблея призывает государства-члены:

10.1. пересмотреть своё антидискриминационное законодательство и при необходимости внести в него поправки, чтобы гарантировать, что оно охватывает все случаи, когда прямая или косвенная дискриминация, включая дискриминацию по ассоциации, может быть вызвана использованием ИИ, и чтобы заявители имели полный доступ к правосудию; в последнем отношении обращать особое внимание на обеспечение презумпции невиновности и обеспечение того, чтобы жертвы дискриминации не сталкивались с непропорциональным бременем доказывания;

10.2. разработать чёткое национальное законодательство, стандарты и процедуры для обеспечения того, чтобы системы на основе ИИ соответствовали правам на равенство и недискриминацию, где бы использование этих прав могло быть затронуто использованием таких систем;

10.3. обеспечить, чтобы органы по вопросам равенства были в полной мере уполномочены решать вопросы равенства и недискриминации, возникающие из-за использования ИИ, и поддерживать лиц, возбуждающих дела в этой области, и чтобы у них были все необходимые ресурсы для выполнения этих задач.

11. Для обеспечения того, чтобы использование технологий на основе искусственного интеллекта государственными властями подлежало адекватному парламентскому надзору и проверке со стороны общественности, Ассамблея призывает национальные парламенты:

11.1. Сделать использование таких технологий частью регулярных парламентских дебатов и обеспечить наличие адекватной структуры для таких дебатов;

11.2. потребовать от правительства уведомить парламент перед развёртыванием такой технологии;

11.3. требуют, чтобы использование таких технологий властями систематически регистрировалось в публичном реестре.

12. Для решения основных проблем разнообразия и инклюзивности в области ИИ Ассамблея также призывает государства-члены:

12.1. способствовать включению женщин, девочек и представителей меньшинств в образовательные программы STEM, с самого раннего возраста и до

самых высоких уровней, и работать вместе с промышленностью для обеспечения разнообразия и интеграции на протяжении всей карьеры;

12.2. поддерживать исследования предвзятости данных и средств, с помощью которых можно эффективно противодействовать ее влиянию в системах на основе ИИ;

12.3. продвигать цифровую грамотность и доступ к цифровым инструментам для всех членов общества.

13. Ассамблея предлагает всем организациям, как государственным, так и частным, которые работают над системами на основе ИИ и с ними, обеспечить, чтобы уважение к равенству и недискриминации было интегрировано с самого начала в разработку таких систем и надлежащим образом протестировано перед их развёртыванием, везде, где эти системы могут повлиять на осуществление основных прав или доступ к ним. С этой целью он предлагает этим организациям рассмотреть вопрос о наращивании потенциала для системы оценки воздействия на права человека для разработки и развёртывания систем искусственного интеллекта как частными, так и государственными организациями. Кроме того, он поощряет использование междисциплинарных и разнообразных команд на всех этапах разработки и развёртывания систем на основе ИИ.

14. Наконец, Ассамблея призывает национальные парламенты поддерживать работу, проводимую на международном уровне, в частности через Специальный комитет Совета Европы по искусственному интеллекту (САНАИ), для обеспечения эффективного применения стандартов прав человека в области ИИ, и что в этой области гарантируется соблюдение принципов равенства и недискриминации.

## **В. Проект рекомендации**

1. Ассамблея ссылается на свою Резолюцию... (2020), озаглавленную «Предотвращение дискриминации, вызванной использованием искусственного интеллекта». Он отмечает, что эта резолюция была принята, поскольку в Совете Европы продолжалась работа Специального комитета по искусственному интеллекту (САНАИ).

2. Ассамблея напоминает, что равенство и недискриминация являются основными правами и что все государства-члены обязаны уважать эти права в соответствии с Европейской конвенцией о правах человека (ETS № 005), как это интерпретируется прецедентным правом Европейского Суд по правам человека и Европейская социальная хартия (ETS No. 035) в интерпретации Европейского комитета по социальным правам.

3. Поэтому Ассамблея призывает Комитет министров принять во внимание особенно серьёзное потенциальное воздействие использования искусственного интеллекта на осуществление прав на равенство и недискриминацию при оценке необходимости и осуществимости международно-правовой основы для искусственного интеллекта. Интеллект.

## **С. Пояснительный меморандум г-на Лакруа, докладчика**

### **Введение**

1. Искусственный интеллект (ИИ) меняет наш образ жизни. Автоматизированные процессы принятия решений используются в процедурах отбора для получения работы или высшего образования; они используются для оценки кредитоспособности человека или для определения его права на получение социальных пособий; они определяют информацию, которая становится доступной для пользователей Интернета в их персонализированных новостных лентах или результатах поиска; они определяют, на кого нацелена политическая и другая реклама и какие сообщения.

2. Часто предполагается, что решения, принимаемые машинами, будут объективными и непредвзятыми. Тем не менее, при этом игнорируется роль, которую обязательно играют люди в разработке действующих алгоритмов, а также предвзятость, уже существующая в данных, используемых для их подачи. Сегодня имеется достаточно доказательств того, что использование ИИ может не только воспроизводить известные дискриминационные результаты, но и создавать новые.

3. Появление искусственного интеллекта, который не подлежит регулированию в рамках суверенного и независимого демократического процесса, таким образом, рискует привести к увеличению нарушений прав человека и, в частности, вызвать, закрепить или даже усугубить дискриминацию и изоляцию, независимо от того, является ли это его явной целью.

4. Вызовы многочисленны и затронут отдельных людей и наше общество в целом. Более того, используемые технологии и алгоритмы не знают границ. Это означает, что национальные меры по предотвращению дискриминации в этой области, хотя и необходимы, сами по себе не могут дать достаточного ответа и делают международное регулирование особенно важным. Самый надёжный способ обеспечить эффективное решение поставленных на карту вопросов прав человека – это использовать сильный, общий, многосторонний подход.

5. В этом отчёте делается попытка определить и предложить базовые международные рамки для ориентированного на человека ИИ, основанные на этических принципах, уважении прав человека и недискриминации, равенстве и солидарности. Общая цель – гарантировать, что права каждого человека гарантированы, в частности, права тех людей, которые наиболее подвержены потенциально дискриминационным последствиям использования ИИ, таких как женщины, этнические, языковые и сексуальные меньшинства, рабочие, потребители, дети, пожилые люди, люди с ограниченными возможностями или другие люди, которым грозит изоляция.

## **Объем отчёта**

6. Широкомасштабное внедрение ИИ затрагивает все больше и больше областей повседневной жизни граждан и может существенно повлиять на их доступ к правам, в том числе вызывая дискриминацию. Поэтому мы, как политики, несём особую ответственность за размышление о возможном регулировании таких систем, в частности, для предотвращения такой дискриминации.

7. С самого начала важно отметить, что отчёт составлялся параллельно с подготовкой рядом комитетов Ассамблеи нескольких других отчётов, касающихся ИИ. Эти отчёты касаются: «Правосудие по алгоритму - роль искусственного интеллекта в полицейской деятельности и системах уголовного правосудия»; «Интерфейс мозг-компьютер: новые права или новые угрозы основным свободам?»; «Правовые аспекты «автономных» транспортных средств»; «Необходимость демократического управления искусственным интеллектом»; «Искусственный интеллект и рынки труда: друг или враг?»; «Искусственный интеллект в здравоохранении: впереди медицинские, юридические и этические проблемы». В контексте этих отчётов возникает много вопросов, касающихся равенства и недискриминации. Хотя я кратко затронул некоторые из этих вопросов в моем собственном отчёте, ради эффективности я намеренно сосредоточил свой анализ на других вопросах, связанных с предотвращением дискриминации, вызванной использованием ИИ.

8. Для целей настоящего отчёта описание концепции ИИ приведено в прилагаемом приложении. Я бы добавил, что в общем смысле термин «искусственный интеллект» расплывчат, а его объем неточен. Однако его суть можно понять по аналогии с человеческим интеллектом. Если последнее можно считать для каждого человека сформированным на основе суммы их опыта и того, что они узнали, ИИ можно понимать как комбинацию «больших данных» (обширные наборы ранее существовавших данных, заменяющие индивидуальный опыт). и использование таких данных в машинном обучении. Примечание.

Последнее включает определение математической модели, основанной на алгоритмах и выполняемой с использованием таких методов, как искусственные нейронные сети, что позволяет машине учиться на заданном наборе данных, чтобы иметь возможность делать (точные) прогнозы при столкновении с неизвестными ситуациями.

9. Это помогает концептуализировать рассматриваемые механизмы. Однако для того, чтобы разработать согласованную структуру регулирования, важно определить порог, за которым система должна регулироваться. С одной стороны, определение ИИ таким образом, чтобы охватить все компьютерное кодирование, означало бы, что каждая программа обработки текста или даже каждый интернет-сайт можно было бы рассматривать как составляющие ИИ, что могло бы быть слишком широким. (Было бы сложно разработать последовательную систему регулирования, применимую ко всем формам ИИ, если бы его определение было слишком широким.) С другой стороны, если используемое юридическое определение основано только на методах и приложениях, которые уже используются, он рискует оказаться непригодным для освещения будущих событий, поскольку законодательные процессы часто протекают медленно, а сектор ИИ развивается необычайно быстро.

10. Я хотел бы подчеркнуть, что когда дело доходит до предотвращения дискриминации, вызванной ИИ, важно не столько то, что такое ИИ или как он работает; наша работа должна быть сосредоточена на системах на основе ИИ в зависимости от того, что они делают. Другими словами, политика и правила, которые мы устанавливаем в отношении ИИ, и, в частности, в отношении предотвращения дискриминации, вызванной использованием ИИ, должны охватывать автоматизированные процессы принятия решений, в частности, когда они основаны на машинном обучении. Запись

11. Цель автоматизированных процессов принятия решений, которые касаются нас в этом отчёте, в целом состоит в том, чтобы делать выбор и / или упорядочивать вещи определенным образом. Какие кандидаты будут приглашены на собеседование и какой уровень оплаты будет предложен отобранному лицу? Кто будет принят на обучение в какой вуз? На какой размер социальных пособий вы имеете право? Какие новости появятся в вашей ленте новостей и в каком порядке?

12. Конечно, любой процесс отбора, независимо от того, автоматизирован он или нет, требует выбора. В настоящем отчёте представляет интерес определение мер, которые национальные власти и другие соответствующие субъекты должны предпринять для обеспечения справедливости результатов, полученных при использовании ИИ в автоматизированных процессах принятия решений, и, в частности, чтобы они не производят и не закрепляют дискриминацию в наших обществах.

## Использование ИИ уже даёт дискриминационные результаты

*«[Приписывание с помощью ИИ более низких кредитных лимитов женщинам] имеет значение для женщины, которая пытается начать свой бизнес в мире, который, кажется, все ещё считает, что женщины не могут быть такими же успешными или кредитоспособными, как мужчины. Для жены, пытающейся выйти из оскорбительных отношений, это важно. Это имеет значение для меньшинств, пострадавших от институциональных предубеждений. Это важно для многих. И это важно для меня». Джейми Хайнемайер Ханссон Примечание.*

13. Как упоминалось выше, автоматизированные процессы принятия решений уже широко используются в повседневной жизни в таких разнообразных областях, как отправление правосудия, процессы отбора для получения высшего образования, найм, «оптимизация» рабочего времени персонала и оценка кредитоспособности или права на получение социальные пособия.

14. Как обсуждается ниже, имеются обширные свидетельства того, что такие процессы часто приводят к несправедливым результатам, дискриминируя по признаку (например) пола, этнического происхождения, социального статуса или психического здоровья. Как законодатели, мы обязаны бороться с такими нарушениями прав человека.

15. В следующих двух разделах, в которых я выделяю некоторые известные случаи дискриминации, уже вызванной использованием ИИ, я проводил различие между частным и государственным использованием ИИ. В этих областях на карту действительно могут быть поставлены различные права, и когда происходит дискриминация, способы возмещения ущерба, потенциально доступные жертвам, также различаются. Затем в третьем разделе я исследую информационные потоки, управляемые с помощью ИИ. Это поднимает разные вопросы и, в частности, может усугубить дискриминацию, усиливая стереотипы и предрассудки.

### Частный сектор

16. ИИ может предоставить мощные инструменты для оптимизации услуг, предоставляемых клиентам, и повышения эффективности бизнеса компании, поскольку он может значительно ускорить процессы, на выполнение которых у людей уйдёт больше времени. Однако использование ИИ также может иметь крайне негативные последствия для некоторых групп людей.

17. Один из крупнейших частных работодателей в мире, Amazon, потратил годы на разработку инструмента подбора персонала на основе ИИ. Он был обучен проверять соискателей с использованием резюме, представленных в компанию за 10-летний период. Однако в 2018 году Amazon решила отказаться от этого инструмента, поскольку он демонстрировал гендерную предвзятость. Примечание. Крупные компании также все чаще используют автоматизированные личностные тесты при найме на работу в качестве средства фильтрации кандидатов. Но было обнаружено, что они дискриминируют кандидатов с психическими расстройствами, исключая их на основании оценок, которые плохо предсказывают эффективность работы.

18. Целевая интернет-реклама, основанная на машинном обучении, – ещё один известный источник дискриминации в сфере занятости. Независимое исследование, например, показало, что значительно меньше женщин, чем мужчин, демонстрируют онлайн-рекламу от компаний, которые помогают в поиске высокооплачиваемой работы. Примечание. Этот вид дискриминации трудно обнаружить, кроме как с помощью обширных исследований, и людям практически невозможно оспорить их. Они не могут знать того, чего не видят.

19. Помимо трудоустройства, Департамент финансовых услуг Нью-Йорка также был приглашён для расследования утверждений о том, что несколько крупных компаний, занимающихся возвратом налогов, использовали рекламные функции Google, чтобы скрыть дополнительные возможности подачи налоговой декларации от лиц с низким доходом, которые имели бы право подавать налоговую декларацию за бесплатно. Международная амнистия выделила методы Facebook, которые позволяют рекламодателям в сфере жилья либо таргетировать, либо исключать определённые группы людей на основании их этнической принадлежности или возраста. Министерство жилищного строительства и городского развития США, кроме того, предъявило Facebook обвинение в марте 2019 года, с помощью своей рекламной платформы, поощряя, разрешая и провоцируя дискриминацию по признаку расы, цвета кожи, религии, пола, семейного положения, национального происхождения и инвалидности. Платформа позволила рекламодателям скрыть свою рекламу от пользователей в определённых районах или отказаться от рекламы жилья пользователям, имеющим определённые интересы, такие как «мода на хиджаб» или «латиноамериканская культура».

20. Примеры, подобные этому, показывают, как поведение в Интернете, такое как выбор интернет-пользователем тем поиска, может использоваться для вывода конфиденциальной частной информации, такой как их этническое происхождение, религиозные убеждения, сексуальная ориентация или гендерная идентичность, или их близость к определенным темам, и таргетировать рекламу способами, которые могут быть дискриминационными. Это также вызывает серьёзные вопросы о защите конфиденциальности.

21. Чтобы соответствовать требованиям закона о борьбе с дискриминацией, Facebook согласился внести изменения в свои алгоритмы, чтобы предотвратить такой таргетинг в будущем. Примечание. Однако алгоритмы на основе машинного обучения, как было показано, вызывают дискриминацию, даже когда рекламодатели не преднамеренно нацелены на рекламу на основе критериев, которые соответствуют характеристикам, защищённым антидискриминационным законодательством. Таким образом, исследование показало, что объявления о вакансиях для низкооплачиваемой работы (дворники, водители такси) показаны большей доле меньшинств, а вакансии учителей и секретарей дошкольных учреждений показаны большей доле женщин.

22. Аналогичные проблемы возникают при использовании систем на основе искусственного интеллекта для оценки кредитоспособности. Например, вскоре после запуска Apple Card в 2019 году сотни клиентов начали жаловаться на сексистские результаты. Мужчинам были предоставлены кредитные лимиты, во много раз превышающие те, которые были предоставлены женщинам в аналогичной ситуации, которые, кроме того, не имели возможности оспорить решение. Единственное объяснение, которое сотрудники Apple смогли дать клиентам, было: «Это просто алгоритм».

23. В моей стране, Бельгии, некоторые медицинские страховые компании в настоящее время стремятся использовать приложения для смартфонов для сбора информации о состоянии здоровья лиц, на которые распространяется действие их страховых полисов. Это вызывает серьёзные проблемы с конфиденциальностью, а также может быть источником дискриминации по состоянию здоровья.

24. Эти примеры – больше, чем просто анекдоты: они ясно показывают, что использование ИИ не только может вызвать дискриминационные эффекты, независимо от основания дискриминации, но и что во многих случаях это уже происходит.

25. Прямая или косвенная такая дискриминация является нарушением основных прав. Если это допустить или не остановить, использование ИИ фактически увековечит дискриминацию, а в некоторых случаях усугубит её. Женщины, ЛГБТИ, люди, принадлежащие к этническим или религиозным меньшинствам, инвалиды и другие, останутся заблокированными на низкооплачиваемой работе на дискриминационных основаниях, с меньшими возможностями доступа к кредитам, товарам и услугам, а также лица, принадлежащие к группам, которые считаются нежелательными с точки зрения жилья. поставщики останутся исключёнными из определённых жилых районов, не только дискриминируя их индивидуально, но и усугубляя пространственную и социальную сегрегацию в обществе.

26. Хотя приведённые выше примеры в основном американские, следует отметить, что вовлечённые компании претендуют на миллионы, а иногда и на миллиарды клиентов на всех континентах, включая Европу, и что их алгоритмы могут иметь дискриминационные последствия в каждом государстве-члене Совета Европы. Парламенты обязаны решать эти проблемы, обеспечивая, чтобы антидискриминационные законы были достаточно жёсткими для защиты людей и борьбы с систематической дискриминацией, чтобы компании, использующие дискриминационный ИИ, могли быть привлечены к ответственности, и чтобы имелись эффективные средства правовой защиты.

## **Государственный сектор**

27. Не только частный сектор прибегает к искусственному интеллекту: он часто используется государственными органами, особенно в контексте государства всеобщего благосостояния, где социальные службы используют цифровые технологии для оценки права людей на получение пособия, расчёта суммы их право или расследование возможного мошенничества или ошибки. Многочисленные примеры вызывают серьёзную озабоченность по поводу цифровых технологий, использующих личные данные таким образом, чтобы нарушить конфиденциальность и / или ошибочно лишить людей социальных льгот.

28. Таким образом, в Нидерландах индикатор риска системы SyRI, используемый для выявления риска мошенничества, собирает и сравнивает данные из нескольких государственных баз данных. В одном из пилотных проектов, запущенных в этом контексте, данные о 63000 человек, которые получали социальные пособия с низким доходом и которые не подозревались в каких-либо правонарушениях, были сопоставлены с данными об использовании воды государственными компаниями, поставляющими воду, для выявления автоматически независимо от того, жили ли люди одни или вместе. Было выявлено 42 случая мошенничества, то есть вероятность успеха составила всего 0,07%. Это вызывает серьёзные вопросы относительно соблюдения права на неприкосновенность частной жизни, а также презумпции невиновности. Голландский суд недавно постановил, что законодательство, регулирующее СИРИ, содержит недостаточную защиту от вмешательства в частную жизнь, поскольку меры, принятые для предотвращения и борьбы с мошенничеством в интересах экономического благополучия, были несоразмерными. Более того, системе не хватало прозрачности, и её нацеливание на бедные районы могло быть равнозначно дискриминации по признаку социально-экономического статуса или статуса мигранта.

29. В других европейских делах Верховный суд Польши отменил систему, созданную в 2014 году, которая разделяла безработных на три категории на основе данных, собранных в момент регистрации для получения пособий и компьютерного собеседования; однако аналогичная система была внедрена в Австрии в 2018 году. В Швеции от автоматизированной системы сбора отчётов о деятельности соискателей работы отказались в 2018 году, поскольку от 10 до 15% автоматизированных решений, принимаемых на основе собранной информации, были признаны неверными.

30. В Соединённом Королевстве универсальная кредитная система, которая объединяет шесть социальных пособий в одно, является первой государственной службой, которая стала цифровой по умолчанию, с алгоритмом, используемым для расчёта пособий каждый месяц на основе информации, получаемой в режиме реального времени от работодателей, налоговые органы и государственные ведомства. Многие люди потеряли льготы, потому что им просто не хватало навыков для заполнения новых онлайн-форм. Более того, выгоды могут автоматически уменьшаться без объяснения причин или уведомления на основе результатов, полученных с помощью алгоритма. Последнее, а не получатель социальных выплат, подвергается сомнению. Однако власти признают, что каждый месяц примерно 2% из миллионов совершаемых транзакций (то есть десятков тысяч случаев) дают неверные результаты. Поскольку пандемия Covid-19 оставляет все больше людей без работы и зависит от социального обеспечения, существует реальный риск того, что большее количество людей будут несправедливо лишены доступа к социальному обеспечению.

31. Аналогичная система, введённая в действие в Австралии (обычно называемая «Robodebt»), дала особенно вредные результаты. Автоматическое сопоставление данных (заменяющее обследование на людях, ранее проводившееся государственными служащими вручную) было введено в 2016 году для выявления расхождений между данными о доходах получателей пособий, хранящимися в налоговых органах и социальных службах, с целью выявления возможных переоплат или мошенничества. С этого момента от любого, для кого алгоритм оценивает несоответствие как подозреваемого, требовалось предоставить доказательства обратного через онлайн-форму, без которой их распределение было бы сокращено или полностью исключено. Алгоритм, однако, взял данные налоговых органов (которые основаны на полном году) и сравнил их с доходом за две недели, игнорируя тот факт, что доход получателей социального обеспечения часто бывает очень нерегулярным, например, из-за краткосрочных контрактов или сезонной работы. В результате тысячи людей были ошибочно лишены социальных выплат, и многие из них не смогли оспорить эти решения (автоматические уведомления отправлялись на старый адрес; у них не было доступа к portalу, через который они могли бы переслать необходимые доказательства). Во многих случаях люди внезапно оказывались в серьёзных долгах, сообщалось о некоторых случаях

суицида. Некоторые источники подсчитали, что власти пытались потребовать от граждан почти 600 миллионов австралийских долларов (360 миллионов евро) на основе этой системы, которая часто приводит к ошибкам, но при которой бремя доказывания перекладывается на отдельного человека, а результаты очень трудно получить вызов.

32. Эти примеры – лишь верхушка айсберга; количество проблем будет расти, поскольку правительства будут стремиться расширить использование технологий во имя большей эффективности. Недавний скандал вокруг автоматической корректировки результатов A-level в Соединённом Королевстве в контексте пандемии Covid-19, которая особенно затронула учеников из неблагополучных районов, и аналогичные проблемы, связанные с результатами международного бакалавриата, – это лишь два таких примера.

33. В отношении недискриминации есть три основных вопроса. Во-первых, как правило, отсутствует предварительная проверка, демократический надзор и общественные дебаты по этим вопросам. Таким образом, парламентские дебаты о введении SyRI в Нидерландах в 2006 г. практически не обсуждались, несмотря на предупреждения органа по защите данных и других сторон. Кроме того, запросы о свободе информации часто нарушаются из-за широких исключений или непонимания властями используемых технологий. Поэтому неравное воздействие таких систем на бедных и маргинализованных слоёв населения часто остаётся незамеченным. Во-вторых, ИИ обычно считается более справедливым и точным, чем люди. Однако, хотя это может быть верно для очень конкретных задач, это гораздо менее определённо там, где необходимо учитывать более широкий контекст. Там, где технологии позволяют масштабировать процессы, но в то же время приводят к крупномасштабным ошибкам, те, кто меньше всего способен бросить вызов системе (например, бедные или пожилые, мигранты, люди с ограниченными возможностями), снова будут непропорционально затронуты. Это особенно серьёзно, когда ИИ используется в контексте государства всеобщего благосостояния. Наконец, цифровые технологии часто преднамеренно нацелены на бедных и маргинализованных людей, что со временем расширяет возможности для постоянного государственного надзора за этими людьми.

34. Сталкиваясь с этими проблемами, парламенты должны понимать, что поставленные на карту вопросы носят не только технический, но и весьма политический характер. Системы на основе искусственного интеллекта дороги в установке, используются для достижения определённых политических целей, которые (в приведённых выше примерах) становятся предметом политизированных дебатов о государстве всеобщего благосостояния и часто обходятся дорого. Поэтому надзор и обсуждение использования этих технологий должны быть частью регулярных парламентских дебатов. Это может быть сделано путём установления правил, например, обязывая правительства уведомлять парламенты заранее об

использовании таких технологий, требуя систематической регистрации их использования в публичном реестре и обеспечивая наличие структуры для таких обсуждений. Парламентариям не обязательно быть экспертами в области искусственного интеллекта, чтобы понимать основные политические и социальные проблемы. Например, в Австралии и Соединённом Королевстве, где автоматизированные расчёты используются в качестве основы для властей, требующих, чтобы граждане возмещали социальные выплаты, во многих случаях ошибочно, бремя доказывания фактически было отменено, и в то же время решения сложно или невозможно оспорить. Не соблюдается надлежащая правовая процедура, и право на средство правовой защиты ограничено. В Нидерландах была нарушена презумпция невиновности и право на неприкосновенность частной жизни десятков тысяч людей, однако было выявлено лишь несколько случаев мошенничества в сфере социального обеспечения. Право на социальную помощь также нарушается, когда автоматизированные процессы затрудняют получение людьми льгот, на которые они имеют право. Больше всего от этого страдают наименее обеспеченные люди в обществе.

### **Частный случай информационных потоков**

«Проблема Интернета в том, что он вознаграждает крайности. Допустим, вы едете по дороге и видите автокатастрофу. Конечно смотришь. Все смотрят. Интернет интерпретирует подобное поведение как то, что все просят об автокатастрофах, поэтому он пытается их удовлетворить». Эван Уильямс, основатель Twitter.

35. Наконец, я хочу привлечь внимание к другой сфере, в которой использование ИИ может усугубить дискриминацию, разжигая экстремизм и ненависть. Было показано, что алгоритмы некоторых веб-сайтов, особенно сайтов социальных сетей, автоматически рекомендуют своим пользователям все более радикальные точки зрения.

36. Президентские выборы в США в 2016 году часто упоминались как случай, когда распространение в Интернете радикальных или крайних взглядов (а также фейковых новостей - вопрос, который, однако, выходит за рамки данного отчёта) могло оказать решающее влияние на исход выборов. Бывший сотрудник Google Гийом Часло создал алгоритм, чтобы выяснить, существует ли предвзятость в видеороликах, автоматически рекомендованных алгоритмом YouTube во время этих выборов зрителям, просмотревшим видеоролики, содержащие слова «Трамп» или «Клинтон». Результаты не только выявили большое количество рекомендаций в отношении видео, выражающих крайние взгляды, но также и то, что независимо от имени кандидата, который изначально искал зритель, алгоритм YouTube с гораздо

большой вероятностью будет рекомендовать видео, которые были более благоприятными для Трампа, чем для Клинтон (часто видеоролики о заговоре против Клинтон).

37. Тенденция алгоритма YouTube рекомендовать экстремальные видео к тому же не ограничивается политической областью: эксперименты в других сферах привели к аналогичным результатам. Например, просмотр видеороликов о вегетарианстве приводит к видеороликам о веганстве; видеоролики о беге трусцой ведут к другим о беге на ультрамарафоне; видеоролики о вакцине против гриппа превращаются в видеоролики с заговором против вакцинации. Аналогичные опасения высказывались и в отношении новостных лент Facebook.

38. Алгоритмы, используемые в социальных сетях и на сайтах газет, как правило, оптимизированы для получения прибыли. Поэтому они по умолчанию продвигают элементы, которые могут привлечь большое количество кликов или «взаимодействий». Таким образом, читателей побуждают проводить на сайте больше времени (и, следовательно, получать больше рекламы), стимулируя их естественную склонность привлекаться к возбуждающим заголовкам.

39. Я хочу подчеркнуть, что эта спираль не является неизбежной точно так же, как ситуация, когда пользователи информационных сайтов и социальных сетей попадают в «пузыри» других пользователей, разделяющих одни и те же точки зрения, не безжалостна. Многие зависит от целей, которые были определены для используемого алгоритма: другими словами, что он был разработан для оптимизации? Цели, отличные от прямой прибыльности, могут быть зафиксированы, например, представление онлайн-читателям как можно более широкого диапазона точек зрения. Этот выбор был сделан некоторыми скандинавскими газетами при разработке алгоритмов, используемых для их интернет-изданий.

## Данные

*«Мусор на входе, мусор на выходе»*

40. Данные играют решающую роль в области ИИ. С одной стороны, огромные наборы данных (так называемые «большие данные») необходимы для обучения и совершенствования машинного обучения с целью разработки сложных систем ИИ. Такие данные обычно касаются людей и часто создаются ими (например, когда они выбирают ссылку в ленте новостей или когда они заполняют онлайн-форму).

41. Несмотря на защиту, введённую, по крайней мере, в странах ЕС, примечанием к Общему регламенту защиты данных (GDPR), пользователи информационных систем, однако, не всегда осведомлены о том, что такие данные собираются или о том, как они могут впоследствии использоваться. Это поднимает множество вопросов в отношении уважения к частной жизни - общая нить, проходящая через многие применения ИИ. Для целей настоящего отчёта я просто подчеркну, что не все пользователи информационных систем обладают одинаковыми знаниями в этой области. Таким образом, не все люди равны в отношении сбора данных в Интернете и защиты их личных данных.

42. С другой стороны, данные никогда не бывают беспристрастными: они всегда зависят от времени и места, где они были собраны. *Примечание. Предвзятость присуща существующим человеческим данным, и как ведёт, так и возникает из стереотипов и предрассудков. Предубеждения, преобладающие в то время и в этом месте, а также в умах тех, кто разрабатывает и проводит сбор данных, отражаются в собранных данных.*

43. Использование предвзятых наборов данных или наборов данных, отражающих исторические предубеждения, предубеждения или дискриминацию, является основной причиной дискриминации в ИИ. Если исторически меньше женщин и / или меньше людей, принадлежащих к этническим меньшинствам, были заняты в определённых областях, или они работали с более низкой заработной платой, или в кредитах было отказано людям, принадлежащим к определенным группам, или меньшинства имели тенденцию взаимодействовать рекламируя аренду дома, а не его покупку, ИИ, который основывает свои решения по оптимизации на распознавании и воспроизведении исторических закономерностей, просто послужит укреплению дискриминации. *Примечание. Для исправления этого недостатка требуется не только знание исторических закономерностей, но и осознанные дизайнерские решения - область, которую я исследую далее ниже.*

44. В некоторых случаях предвзятость можно легко исправить - например, когда дело доходит до программного обеспечения для распознавания лиц, которое работает менее точно в зависимости от цвета кожи, использование более широкого диапазона фотографий для обучения машины может исправить некоторые проблемы. Расширение набора данных в таком случае может быть относительно простым, поскольку в Интернете бесплатно доступен широкий спектр фотографий людей различного этнического происхождения. Тем не менее, насколько легко решить такую проблему, также зависит от того, как используются данные. Обратите внимание, что в печально известном случае с Google Фото, когда приложение, которое изначально было выпущено, автоматически помечало фотографии афроамериканцев как «гориллы», расистскую маркировку можно было относительно легко исправить. Когда методы распознавания лиц на основе искусственного интеллекта используются не просто для классификации

изображений как принадлежащих к определенным группам, но и в системе уголовного правосудия, например, для идентификации конкретных лиц в ситуациях, когда их личная свобода может быть поставлена на карту и где такие системы работают значительно хуже именно для людей с тёмной кожей могут потребоваться гораздо более сложные решения, чтобы устранить далеко идущие последствия таких недостатков для прав человека.

45. Некоторые данные исторически не измерялись вообще, что привело к тому, что определённые группы людей были невидимы в доступных наборах данных. Люди, которые не идентифицируют себя ни как мужчина, ни как женщина (в частности, некоторые интерсекс-люди), не только испытывают враждебные онлайн-формы, где от них требуется отметить поле «мужской» или «женский», но их конкретная ситуация не может быть измерена и подвергнута дискриминации. Их нельзя ни идентифицировать, ни предотвратить. В области медицинского тестирования женщины также исторически исключались из медицинских исследований, что означает, что их здоровье и реакция организма на лекарства менее понятны, чем у мужчин. Подобные проблемы предшествуют использованию систем на основе ИИ, но они означают, что использование таких систем, обученных с использованием исторических данных, будет иметь тенденцию воспроизводить и укреплять существующую дискриминацию.

46. Более того, данные дают определённое представление о реальности, но часто упрощают ее. В результате они могут лишь дать более или менее приблизительное представление о реальности, которую они должны представлять. Примечание. Многие государства, например, отказываются разрешить сбор этнических данных (часто аргументируя это тем, что неправильное использование таких данных в прошлом показывает, что они должны больше никогда не будет собираться). Вместо этого используются прокси, такие как страна рождения человека, его родители или их бабушки и дедушки. Эти прокси-серверы охватывают многих людей, принадлежащих к этническим меньшинствам, но упускают из виду многих других, чьи семьи в течение нескольких поколений жили в стране и которые все ещё могут сталкиваться с дискриминацией по признаку цвета кожи или языка.

47. В других случаях информацию, относящуюся к алгоритму, может быть чрезвычайно сложно вычислить. Это особенно касается абстрактных понятий (например, справедливости - ключевой вопрос в судебной сфере), в отличие от количественных элементов, таких как количество преступлений с применением ножа, зарегистрированных в данной области за определённый период времени. Вопрос о справедливости по алгоритму является предметом постоянной работы другого комитета, и поэтому я не буду рассматривать его подробно в этом отчёте. Однако я хочу подчеркнуть серьёзную дискриминацию, которая может возникнуть, когда алгоритмы отдадут приоритет «эффективности» (опираясь на элементы, которые легко подсчитать), а не справедливости (с учётом более широких

последствий результатов алгоритма для общества в целом). Печально известная программа COMPAS, используемая в некоторых американских штатах для анализа риска рецидивизма (чтобы помочь судьям принять решение о вынесении или не назначении тюремного заключения), представляет собой достойный сожаления пример.

48. Таким образом, в целом способность систем на основе ИИ предотвращать дискриминацию во многом зависит от данных, используемых для их обучения. Историческая предвзятость, отсутствие данных по ключевым вопросам, использование прокси-серверов и трудности, связанные с количественной оценкой абстрактных концепций, - все это необходимо решать и эффективно решать при развёртывании систем на основе ИИ, где бы их использование могло повлиять на права человека. Как законодатели, мы должны осознавать важнейшие проблемы прав человека, стоящие на кону в этой области, и разрабатывать способы обеспечения защиты граждан от такой дискриминации.

## **Дизайн и назначение**

«Царь Мидас сказал: «Я хочу, чтобы все, к чему я прикасалась, превратилось в золото», и он получил именно то, о чем просил. Это была цель, которую он поместил в машину, так сказать, а затем его еда, его питье и его родственники превратились в золото, и он умер в страданиях и голоде». Стюарт Рассел, исследователь ИИ.

49. Алгоритмы предназначены для целеустремлённой работы для достижения конкретной цели, поставленной их программистами. Математическая модель разработана, чтобы позволить машине учиться на заданном наборе данных, и оптимизирована, чтобы иметь возможность делать точные прогнозы при столкновении с неизвестными ситуациями. Выбор подходящей цели для алгоритма имеет решающее значение, поскольку все будущие проектные решения и результаты, в конечном итоге получаемые с использованием системы на основе ИИ, будут зависеть от этого выбора. Непродуманные цели или выбор политики приведут к нежелательным и несправедливым результатам.

50. Автоматизированные процессы найма, используемые, например, крупными компаниями, часто оптимизируются с точки зрения «эффективности», что означает, что они служат для максимального сокращения числа кандидатов, которые должны пройти собеседование или чтобы их заявки просматривались людьми. Однако, если алгоритм предназначен для отбора кандидатов, которые кажутся подходящими для существующей корпоративной культуры, он вряд ли поможет увеличить разнообразие внутри этой компании или изменить её культуру.

Обратите внимание, что соображения «эффективности» также лежат в основе многих проблемных государственных секторов. использование ИИ, в частности, в контексте государства всеобщего благосостояния, как я описал ранее.

51. Если цель модели машинного обучения несовместима с необходимостью избегать дискриминации, то результаты, которые она даёт, снова увековечивают или усугубляют дискриминацию. Рекламный инструмент Facebook, о котором говорилось ранее, предлагал рекламодателям на выбор ряд целей оптимизации: количество просмотров рекламы, степень вовлеченности, которую она произвела, или объем продаж, к которым она привела. Если бы показ более высокой доли белых пользователей домов для покупки приводил к большему вовлечению, то алгоритм делал бы это, тем самым дискриминируя чернокожих пользователей. Это произошло потому, что алгоритм был оптимизирован для бизнес-целей, без учёта необходимости обеспечения уважения права человека, такие как равный доступ к жилью.

52. В случае информационных потоков также остаётся серьёзной проблемой тот факт, что алгоритмы предназначены для оптимизации взаимодействия. «Бесплатные» платформы, такие как Facebook и YouTube, а также онлайн-СМИ, финансируемые за счёт рекламы, заинтересованы в увеличении времени, которое пользователи проводят на своих сайтах, и, таким образом, в увеличении количества рекламных объявлений. Если данные показывают, что пользователей больше всего привлекает экстремальный контент, алгоритмы будут продвигать этот контент. Хотя такие платформы склонны утверждать, что они «просто служат зрителям то, что они хотят», цель алгоритмов - максимизировать доходы от рекламы для соответствующих компаний. Примечание. В процессе игнорируется вопрос о том, что действительно является информативным или заслуживающим внимания, а часто пропагандируются сексистские, расистские, антисемитские, исламофобские, антицыганские, гомофобные, трансфобные и другие высказывания, разжигающие ненависть.

53. Ещё до того, как будет определена конкретная цель, для которой должен быть оптимизирован алгоритм, необходимо задать фундаментальные вопросы о назначении системы на основе ИИ. Чтобы взять пример из «реального» мира, программы останова и поиска будут усиливать предвзятость в системе уголовного правосудия, когда они нацелены только на бедные районы, ищущие определённые типы преступного поведения, оставляя в покое более богатые районы (где преступления белых воротничков часто происходят, и домашнее насилие может произойти так же, как и в других местах). Данные, собранные с помощью предвзятых стратегий полицейской деятельности и введённые в системы на основе ИИ, автоматически покажут более высокий уровень преступной активности в целевых областях, что затем приведёт к усилению полицейской деятельности в этих областях, создавая глубоко дискриминационную «петлю обратной связи», которая

наносит ущерб жители более бедных кварталов (часто представители этнических меньшинств), не в состоянии зафиксировать активность в других районах, которая в равной степени заслуживает внимания полиции. Такие «петли обратной связи» возникают из-за того, что алгоритмические результаты, основанные на искажённых наборах данных, как правило, подтверждают дискриминационные методы – и потому, что не проводится оценка того, чего не хватает в этих наборах данных.

54. Следует подчеркнуть, что – как и любая процедура выбора – ни один автоматизированный процесс принятия решений никогда не будет полностью нейтральным, поскольку он всегда будет результатом выбора дизайна, а именно конкретных концепций того, как следует упорядочивать вещи. Однако из-за их способности принимать большое количество решений с очень высокой скоростью последствия внедрения дискриминационных систем на основе ИИ могут быть драматичными. Как законодатели, мы должны найти способы обеспечить, чтобы выбор дизайна, сделанный при разработке и внедрении автоматизированных процессов принятия решений, систематически учитывал необходимость защиты прав человека и предотвращения дискриминации.

## **Разнообразие**

«Чернокожие девушки должны научиться программировать» – это оправдание того, что они не решают проблему постоянной маргинализации черных женщин в Кремниевой долине» Сафия Умоджа NobleNote

55. Алгоритмы отражают ценности, убеждения и убеждения тех, кто их разрабатывает. Способность автоматизированных процессов принятия решений учитывать и отражать разнообразие, присутствующее в наших обществах, также зависит от этих факторов.

56. Отсутствие разнообразия, которое обычно существует в технологических компаниях, которым поручено проектировать алгоритмы, вызывает здесь серьёзные проблемы. Недопредставленность женщин в сфере науки, техники, инженерии и математики (STEM), и профессий уже была признана Ассамблеей, которая предложила государствам принять меры для поощрения женщин и девочек к получению школьного и высшего образования в этих областях. Примечание. Однако проблемы выходят за рамки гендерного неравенства. Исследователь Кейт Кроуфорд выделила социальную, расовую и гендерную эндогамию, которая характеризует среду, в которой сегодня нанимают тех, кто тренирует искусственный интеллект.

57. Однако разнообразие в рабочей силе – это не просто проблема «конвейера». Дискриминация на рабочем месте также приводит к высокой текучести кадров среди женщин и меньшинств, которым удалось к нему

присоединиться. «Братские» и «братские» культуры, сексуальные домогательства и гендерная предвзятость на рабочем месте способствуют тому, что женщины в технологических компаниях оставляют свою карьеру на середине в два раза чаще, чем мужчины. Сохраняющееся расистское восприятие цветных людей как «нетехнических» ведёт к их маргинализации и в технической сфере.

58. Отсутствие разнообразия в технологических компаниях не просто дискриминационное само по себе: оно прямо выражается в дискриминационном ИИ. Сафия Умоджа Ноубл в своей работе над поисковыми системами глубоко исследовала, как доминирующие группы могут классифицировать и организовывать представления других таким образом, чтобы воспроизводить и усугублять расистские предрассудки. Примечание. Дизайн онлайн-форм также имеет тенденцию соответствовать только простейшая из жизней, за исключением тех, кто не соответствует норме. Примечание. Повсеместное использование женских персонажей для чат-ботов, которые часто играют подчинённую роль, также увековечивает вредные гендерные стереотипы.

59. Для решения этих проблем жизненно важно - но недостаточно - увеличивать разнообразие рабочей силы, занятой разработкой ИИ, путём принятия решительных мер по расширению доступа женщин и меньшинств к профессиям STEM. Также крайне важно обеспечить, чтобы все студенты, изучающие эти предметы, разрабатывающие технологии для людей, прошли подготовку и получили образование в области истории маргинализированных людей. Применение междисциплинарного подхода к разработке ИИ, с привлечением с самого начала не только технических экспертов, но и экспертов в области социальных, гуманитарных и юридических наук, также во многом поможет предотвратить дискриминацию, вызванную использованием ИИ.

### **Синдром «чёрного ящика»: прозрачность, объяснимость и подотчётность**

*«Мне не дали никаких объяснений. Невозможно доказать свою правоту». Джейми Хайнмайер Ханссон Примечание*

60. Обычный опыт жертв дискриминации, вызванной использованием ИИ, заключается в том, что даже если они могут доказать, что дискриминация имела место, они не могут получить никаких объяснений относительно того, почему был получен дискриминационный результат. Французские студенты, например, столкнулись с такими ситуациями, когда в 2018 году заявки в вузы начали рассматриваться в рамках новой онлайн-системы Parcoursup. Примечание. Часто, как в случае с Apple Card, о котором говорилось ранее, разработчики алгоритма

*сами не могут объяснить почему именно он дал данный результат или набор результатов. Конкретные элементы, которые учитываются конкретным алгоритмом, и вес, придаваемый каждому элементу, редко известны – и частные компании, вложившие средства в разработку таких алгоритмов, обычно неохотно раскрывают их, поскольку они рассматривают их как свои ценная) интеллектуальная собственность.*

61. Это отсутствие прозрачности и объяснимости часто называют синдромом «чёрного ящика». Он может принимать решения, достигаемые с помощью ИИ, которые чрезвычайно трудно оспорить людям. Это также создаёт препятствия для национальных органов по вопросам равенства, которые стремятся поддержать истцов в возбуждении дел в этой области. Однако использование ИИ может иметь серьёзные пагубные последствия для некоторых групп или отдельных лиц, прямо или косвенно дискриминируя их, и может усугублять предрассудки и маргинализацию. Крайне важно обеспечить наличие эффективных и доступных средств правовой защиты для борьбы с дискриминацией, когда она имеет место, и чтобы виновные в такой дискриминации могли нести за неё ответственность.

62. В следующей главе я очерчиваю некоторые элементы, которые могут быть включены в правовые нормы, чтобы гарантировать, что они обеспечат надёжную основу как для предотвращения дискриминации, вызванной использованием ИИ, так и для рассмотрения случаев, когда они возникают. Здесь я хочу подчеркнуть, что, когда компании или государственные органы сталкиваются с доказательствами дискриминации, они действительно находят способы решить проблему – будь то из-за общественного давления или для исполнения решения суда. Наша задача – обеспечить, чтобы такие меры были приняты раньше, чем позже.

### **Необходимость в общем наборе правовых стандартов, основанных на согласованных этических принципах и принципах прав человека**

63. Алгоритмическая предвзятость – это не чисто техническая проблема для исследователей и технических специалистов, но поднимает вопросы прав человека, которые волнуют всех нас и выходят за рамки (необязательных) этических хартий, которые уже были составлены крупными компаниями. Мы не можем позволить себе прятаться за сложностями ИИ в форме усвоенной беспомощности, которая мешает нашим обществам, и в особенности нам как законодателям, вводить правила, призванные защищать и продвигать равенство и недискриминацию в этой области. Примечание. Нам срочно необходимы процедуры, инструменты и методы для регулирования и аудита этих систем, чтобы обеспечить их соответствие международным стандартам в области прав человека. Безусловно, необходимо эффективное внутреннее законодательство, но также и международные стандарты,

учитывая сильные транснациональные и международные аспекты технологий на основе ИИ.

64. Не стремясь быть исчерпывающим, я хотел бы обратить внимание здесь на некоторые ключевые тексты, которые уже были приняты, а также на некоторую работу, которая ведётся на международном уровне и которая помогает подготовить путь вперёд. В этом отношении я с интересом отмечаю начало работы Специального комитета по искусственному интеллекту (САНАИ) 18 ноября 2019 года. Этому межправительственному комитету Совета Европы поручено, в частности, изучить осуществимость и потенциальные элементы по основы широких консультаций с участием многих заинтересованных сторон, правовой основы для разработки, разработки и применения ИИ, основанной на стандартах Совета Европы в области прав человека, демократии и верховенства закона. Примечание. Я также с интересом отмечаю рекомендацию Комиссар Совета Европы по правам человека, «Распаковка искусственного интеллекта: 10 шагов по защите прав человека», в котором чётко отмечается важность вопросов равенства и недискриминации в этой области. Я также отмечаю исследование, озаглавленное «Дискриминация, искусственный интеллект и алгоритмическое принятие решений», которое проливает ценный свет на риски дискриминации, вызванные процессами алгоритмического принятия решений и другими типами ИИ. Примечание. Наконец, я также приветствую вклад Рекомендации Rec / CM (2020) 1 Комитета министров о воздействии алгоритмических систем на права человека, сопровождаемый руководящими принципами по этому вопросу.

65. Помимо работы, проводимой Советом Европы, заслуживают упоминания многие важные инициативы на европейском и международном уровне, в том числе Руководство по этике для заслуживающего доверия ИИ, опубликованное Европейской комиссией в апреле 2019 года, записка и исследование, опубликованное ее Фондом фундаментальных исследований. Агентство по правам человека в области качества данных и искусственного интеллекта – снижение предвзятости и ошибок; Обратите внимание на Принципы ОЭСР в отношении ИИ, принятые в мае 2019 года; Обратите внимание на декларацию G20 об искусственном интеллекте, ориентированном на человека, принятую в июне 2019 года; Примечание и результаты работы нескольких стран, включая ЮНЕСКО, который все ещё продолжался на момент подготовки настоящего доклада.

66. Основываясь на анализе такой ключевой работы на международном уровне, в приложении к настоящему отчёту выделяются пять основных этических принципов, которые должны лежать в основе всей нормативной работы в области ИИ: прозрачность, справедливость и справедливость, ответственность (или подотчётность), безопасность и защищенность, и конфиденциальность. Он подчёркивает, что степень, в которой соблюдение этих основных принципов должно быть встроено в конкретные системы искусственного интеллекта, зависит от

предполагаемого и прогнозируемого использования этих систем. По сути, чем больше потенциальный вред, который может быть причинен использованием данной системы ИИ, тем более строгие требования должны соблюдаться. Здесь я хотел бы подчеркнуть, что право на равенство и недискриминацию является фундаментальным и неотъемлемым элементом демократии, основанной на правах человека и законах. Его необходимо неукоснительно соблюдать и защищать, и нельзя допускать, чтобы использование ИИ подрывало эти принципы, независимо от контекста.

67. В конкретном контексте равенства и недискриминации есть некоторые моменты, по которым мы должны проявлять особую бдительность. Во-первых, все основания дискриминации должны быть эффективно охвачены антидискриминационным законодательством, включая, помимо прочего, реальный или предполагаемый пол, пол, возраст, национальное или этническое происхождение, цвет кожи, язык, религиозные убеждения, сексуальную ориентацию, гендерную идентичность, половые характеристики, социальное происхождение, гражданское положение, инвалидность, состояние здоровья и так далее. Перечень охраняемых территорий, установленный законом, должен быть открытым – то есть, хотя он должен быть как можно более полным, его не следует определять как исчерпывающий список.

68. Во-вторых, необходимо эффективно учитывать как прямую, так и косвенную дискриминацию. Это особенно важно, если в некоторых странах не разрешён сбор определённых данных (например, этнических данных), но алгоритмы определяют такие характеристики на основе прокси (страна рождения человека или его родителей, почтовый индекс, интересы поиска и т. д.). По аналогичным причинам закон должен также запретить дискриминацию на основе реальной или предполагаемой связи человека с определённой группой.

69. В-третьих, учитывая особые трудности, связанные с доказательством дискриминации, вызванной использованием ИИ, крайне важно обеспечить, чтобы жертвы дискриминации не сталкивались с чрезмерным бременем доказывания. Как я уже отмечал ранее в этом отчете, последствия для прав человека требования продемонстрировать свою «невиновность» перед лицом автоматизированного принятия решений могут быть огромными. Совместное бремя доказывания, установленное Директивами ЕС о равенстве, служит здесь полезной моделью.

70. В-четвёртых, очень важны эффективные механизмы правоприменения, и здесь решающее значение может иметь поддержка национальных органов по вопросам равенства. Многие уже начали критически относиться к той роли, которую они могут сыграть, чтобы гарантировать и продвигать равенство в мире, где использование ИИ постоянно расширяется. Примечание: мы должны поддерживать эту работу и обеспечивать, чтобы органы по вопросам равенства располагали адекватными ресурсами для её эффективного выполнения.

71. Наконец, прозрачные бизнес-модели особенно важны там, где системы на основе искусственного интеллекта влияют на выбор, предлагаемый физическим лицам в сети, и когда людям сложно сравнивать алгоритмические результаты. Перед развёртыванием искусственный интеллект также должен подвергаться тщательному тестированию на предвзятость, а затем – не менее тщательному, систематическому и периодическому тестированию предвзятости, и он может вести себя так, как не ожидали разработчики с самого начала.

72. Вышеупомянутые соображения конкретно связаны с проблемами дискриминации, вызванной использованием ИИ. В более широком смысле, я хочу подчеркнуть важность вовлечения гражданского общества и граждан с самого начала в любые размышления по этим вопросам, чтобы обеспечить критическое участие и совместное принятие алгоритмов. Кроме того, регулирование алгоритмов могло бы идти рука об руку с созданием алгоритмических альтернатив, предлагаемых государственными органами (посредством действительно активной политики в этой области), чтобы уравновесить чисто коммерческую логику крупных технологических компаний; эта технологическая политика, основанная на общественной и этической инициативе, будет реализована в дополнение к регулирующей политике.

Что касается признания прав человека, то прецедентная практика Европейского суда по правам человека должна, как всегда, служить ориентиром для государств при разработке внутренних правовых стандартов и устанавливать «красные линии» для их действий. Наконец, учитывая, что ИИ – это быстро развивающаяся область, было бы полезно заранее изучить возможность признания новых прав человека, таких как право на автономию человека, право на прозрачность и оправдание, право на обзор AI, или право на моральную неприкосновенность.

## **Выводы**

74. Как и в случае с большинством технологических разработок, искусственный интеллект не является ни хорошим, ни плохим по своей сути: как острый нож, который можно использовать как в доброкачественных, так и в очень вредных целях, проблема не обязательно в инструменте, а в том, как он разработан и (некоторые) способы его использования. Этот отчёт твердо направлен на то, чтобы ИИ был не просто умным, но полезным для людей и нашего общества.

75. Использование ИИ может повлиять на ряд прав (защита личных данных, частной жизни, доступ к работе, доступ к государственным услугам и социальному обеспечению, стоимость страхования, доступ к правосудию, право на справедливое судебное разбирательство, бремя доказательство, и может затронуть одни группы

больше, чем другие. Хотя это не обязательно создаёт неравенство, оно рискует усугубить уже существующие в обществе и привести к несправедливым результатам во многих сферах жизни.

76. Машинное обучение основано на использовании огромных наборов данных; но данные всегда предвзяты, поскольку они отражают дискриминацию, уже имеющуюся в обществе, а также предвзятость тех, кто собирает и анализирует данные.

77. Алгоритмы создаются людьми, и очень часто однородными командами. Отсутствие разнообразия в мире ИИ создаёт контекст, способствующий воспроизведению разрушительных стереотипов. Диверсификация команд, работающих в этой области, является серьёзной проблемой, которая требует работы не только на начальном этапе, в области образования, но и для обеспечения того, чтобы рабочая среда была инклюзивной в долгосрочной перспективе.

78. Отсутствие прозрачности систем ИИ часто скрывает их дискриминационные эффекты, которые наблюдаются только после того, как система была развёрнута, и многие люди подверглись её негативным последствиям. Чтобы избежать таких последствий, дизайнеры должны с самого начала интегрировать плюралистическую, междисциплинарную и инклюзивную перспективу: кто будет использовать (или на что повлияет?) Эту систему? Будет ли это давать справедливые результаты для всех? Если нет, что мне нужно исправить, чтобы исправить несправедливые результаты? Как можно определить цели инструмента (результаты, для которых его следует оптимизировать), чтобы гарантировать, что они не приводят к дискриминационным результатам?

79. Оценка того, в какой степени соблюдаются принципы равенства и недискриминации, до развёртывания инструмента искусственного интеллекта тем более важна, что многие права человека могут оказаться под угрозой. Кроме того, может быть чрезвычайно сложно оспорить результаты алгоритма из-за его эффекта «чёрного ящика». Тем не менее, чаще всего страдают именно те, кто наиболее маргинализирован и меньше всего может оспорить результаты, иногда с разрушительными последствиями.

80. государствам должны ответить на эти вопросы и решить проблемы, которые они ставят перед нашим обществом сейчас и в будущем. Проблемы выходят за рамки границ и, следовательно, также требуют транснациональных ответов. Частные лица в этой области, которые являются основными разработчиками инструментов на основе ИИ, также в большинстве своём принимают непосредственное участие в поиске решений. Парламенты также должны проявлять большой интерес к последствиям растущего использования ИИ и его влиянию на граждан. Вместе нам необходимо регулировать эти системы, чтобы ограничить их

дискриминационное воздействие и обеспечить эффективное средство правовой защиты в случае дискриминации.

81. В сознании людей искусственный интеллект часто связан с инновациями, но он также должен, что более важно, отвечать другому императиву: инклюзивности. Цель состоит не в том, чтобы препятствовать инновациям, а в том, чтобы регулировать их таким образом, который соразмерен рассматриваемым вопросам, чтобы системы ИИ полностью включали уважение к равенству и запрещение всех форм дискриминации.

## **Необходимость демократического управления искусственным интеллектом**

Отчет | Док. 15150 | 24 сентября 2020 г.

Комитет Комитет по политическим вопросам и демократии

Докладчик: Г-жа Дебора БЕРГАМИНИ, Италия, EPP / CD

Происхождение Ссылка на комитет: Doc 14868, Ref. 4445 от 12 апреля 2019 г. 2020 г. – октябрь Постоянный комитет

### **Резюме**

Искусственный интеллект (ИИ) – это часть новой реальности. Его широкое использование будет во все большей степени влиять на различные аспекты нашей жизни и преобразовывать наши общества. Это влияние может быть как полезным, так и вредным.

В отчете основное внимание уделяется влиянию ИИ на демократию. В нем представлен обзор различных способов, которыми использование технологий на основе ИИ может и уже влияет на функционирование демократических институтов и процессов, а также на социальное и политическое поведение граждан. В нем делается вывод о том, что использование ИИ и его потенциал для злоупотреблений со стороны государств и государственных агентств, а также крупных частных корпораций представляет реальную угрозу для институтов, процессов и норм наших демократических государств, основанных на правах человека.

Необходима структура, обеспечивающая разработку и использование этой технологии при полном уважении наших ценностей, основных прав, верховенства закона и демократии. Совет Европы должен играть ведущую роль в разработке

способов и форматов для обеспечения того, чтобы технологии на основе ИИ использовались для улучшения, а не для нанесения ущерба демократии.

В проекте резолюции приветствуется создание Специального комитета по искусственному интеллекту (САНАИ) и содержится призыв к государствам-членам совместно работать над юридически обязательным документом, направленным на обеспечение демократического управления ИИ, и, при необходимости, дополнять его отраслевыми правовыми инструментами. В проекте рекомендации Комитету министров предлагается выразить поддержку разработке такого документа, возможно, в форме конвенции.

## **А. Проект резолюции Примечание**

1. Технологии всегда оказывали сильное влияние на ход истории человечества. Тем не менее, темпы технического прогресса никогда не были такими быстрыми, а его влияние на людей так же прямое, ощутимое и широкомасштабное, как сейчас, на пороге четвертой промышленной революции. Искусственный интеллект (ИИ), который является его ключевой движущей силой, в целом считается определяющим фактором для будущего человечества, поскольку он существенно изменит жизнь людей и повлияет на человеческие сообщества.

2. Устройства на базе искусственного интеллекта уже широко используются в нашей повседневной жизни и выполняют множество задач, которые ранее выполнялись людьми, как в личном, так и в служебном качестве. Алгоритмы прогнозирования, присущие ИИ, часто используются для принятия важных решений, таких как поступление в университеты, решения по кредитам и управление человеческими ресурсами, а также для пограничного контроля (в том числе в аэропортах) и предотвращения преступности (с помощью методов прогнозирования и использования в преступной среде). системы правосудия, инструментов оценки риска при повторном правонарушении). Поскольку все наши общества изо всех сил пытаются бороться с продолжающейся пандемией Covid-19, ИИ также используется для улучшения фармацевтических исследований и помощи в анализе медицинских данных.

3. Однако долгосрочное влияние ИИ на людей и общество все еще далеко не ясно. Хотя ИИ может открывать большие возможности для продвижения экономического и социального прогресса, он также создает ряд сложных проблем. С одной стороны, ожидается, что ИИ приведет к значительному увеличению производительности и экономического роста, научным открытиям, улучшению здравоохранения, увеличению продолжительности жизни, безопасности и постоянно растущему удобству. С другой стороны, есть опасения, что ИИ может серьезно

подорвать рынки труда по всему миру, привести к увеличению доходов, благосостояния и социальному неравенству, а также поставить под угрозу социальную и политическую стабильность, а также международную безопасность.

4. Технологии на основе ИИ влияют на функционирование демократических институтов и процессов, а также на социальное и политическое поведение граждан. Его использование может оказать как положительное, так и разрушительное воздействие на демократию. Действительно, быстрая интеграция технологий искусственного интеллекта в современные средства коммуникации и платформы социальных сетей предоставляет уникальные возможности для целевого, персонализированного и часто незамеченного влияния на отдельных лиц и социальные группы, которое различные политические субъекты могут соблазнить использовать в своих интересах.

5. С положительной стороны, ИИ можно использовать для повышения подотчетности и прозрачности правительства, помощи в борьбе с коррупцией и получения множества преимуществ для демократических действий, участия и плюрализма, делая демократию более прямой, эффективной и отвечающей потребностям граждан. Технологии на основе ИИ могут расширить пространство для демократического представительства за счет децентрализации информационных систем и коммуникационных платформ. Искусственный интеллект может усилить информационную автономию граждан, улучшить способы сбора информации о политических процессах и помочь им дистанционно участвовать в этих процессах, облегчая политическое выражение и обеспечивая каналы обратной связи с политическими деятелями. Это также может способствовать укреплению доверия между государством и обществом и между самими гражданами.

6. Однако ИИ может использоваться – и, как сообщается, – используется для подрыва демократии путем вмешательства в избирательные процессы, персонализированного политического таргетинга, формирования поведения избирателей и манипулирования общественным мнением. Кроме того, похоже, что ИИ использовался для усиления распространения дезинформации, «эхо-камер», пропаганды и разжигания ненависти, подрыва критического мышления, способствуя росту популизма и поляризации демократических обществ.

7. Более того, широкое использование государствами и частными субъектами технологий на основе искусственного интеллекта для контроля над людьми, таких как автоматическая фильтрация информации, составляющая цензуру, массовое наблюдение с использованием смартфонов, сбор личных данных и отслеживание своей активности в сети и офлайн, может привести к подрыв психологической целостности граждан, гражданских прав и политических свобод и появление цифрового авторитаризма – нового социального порядка, конкурирующего с демократией.

8. Концентрация данных, информации, власти и влияния в руках нескольких крупных частных субъектов, участвующих в разработке и предоставлении технологий и услуг на основе ИИ, а также растущая зависимость отдельных лиц, учреждений и общества в целом от этих услуг также являются основанием для беспокойства. Эти крупные компании больше не служат простыми каналами связи между людьми и учреждениями, а сами по себе играют все более заметную роль, контролируя и фильтруя информационные потоки, осуществляя автоматическую цензуру контента, публикуемого в социальных сетях, устанавливая повестку дня, формируя и трансформируя социальные и социальные сети. политические модели. Действуя на основе бизнес-моделей, в которых прибыль акционеров ставится выше общего блага, эти субъекты могут представлять угрозу демократическому порядку и не должны оставаться за пределами демократического контроля.

9. Ассамблея отмечает, что в последние годы правительства, гражданское общество, международные институты и компании вели обширные дискуссии с целью определения набора общепринятых принципов того, как реагировать на озабоченности, связанные с использованием ИИ. Он приветствует тот факт, что Совет Европы, как ведущая правозащитная организация, активно участвовал в этих обсуждениях будущего ИИ и его управления, и в частности, вклад в этот процесс Комитета министров, Комиссара по Права человека и органы межправительственного сотрудничества.

10. Ассамблея считает, что этические принципы и политика саморегулирования, добровольно вводимые частными субъектами, не являются адекватными и достаточными инструментами для регулирования ИИ, поскольку они не обязательно ведут к демократическому надзору и подотчетности. Европе необходимо обеспечить, чтобы возможности ИИ регулировались и использовались для общего блага.

11. Таким образом, Ассамблея твердо убеждена в необходимости создания сквозной нормативной базы для ИИ с конкретными принципами, основанными на защите прав человека, демократии и верховенства закона. Любая работа в этой области должна включать все заинтересованные стороны, включая, в частности, граждан и крупные частные компании, участвующие в разработке и предоставлении технологий и услуг на основе ИИ.

12. Совет Европы, как ведущая международная организация, устанавливающая стандарты в области демократии, должен играть ведущую роль в разработке способов и форматов, обеспечивающих использование технологий на основе ИИ для улучшения, а не для нанесения ущерба демократии.

13. В этом контексте он приветствует создание Комитетом министров Специального комитета по искусственному интеллекту (САНАИ) для изучения осуществимости и потенциальных элементов на основе широких консультаций с

участием многих заинтересованных сторон правовой основы для дизайн, разработка и применение искусственного интеллекта. Он призывает государства-члены Совета Европы и другие государства-наблюдатели, участвующие в САНАИ, совместно работать над юридически обязательным инструментом, направленным на обеспечение демократического управления AI и, при необходимости, дополнить его отраслевыми правовыми инструментами.

14. Ассамблея считает, что такой документ должен:

14.1. гарантировать, что технологии на основе искусственного интеллекта разрабатываются, разрабатываются и эксплуатируются в полном соответствии со стандартами Совета Европы в области прав человека, демократии и верховенства закона и в их поддержку;

14.2. способствовать общему пониманию и обеспечивать соблюдение ключевых этических принципов и концепций и выполнение вышеупомянутых стандартов, включая:

14.2.1. прозрачность, в том числе доступность и объяснимость;

14.2.2. справедливость и беспристрастность, включая недискриминацию;

14.2.3. человеческая ответственность за решения, включая ответственность и доступность средств правовой защиты;

14.2.4. безопасность и защита;

14.2.5. конфиденциальность и защита данных;

14.3. стремиться максимально увеличить возможное положительное влияние ИИ на функционирование демократических институтов и процессов, включая, в частности:

14.3.1. улучшить подотчетность правительства;

14.3.2. Помогают бороться с коррупцией и экономической преступностью;

14.3.3. способствовать демократическим действиям, участию и плюрализму;

14.3.4. сделать демократию более прямой, эффективной и отвечающей потребностям граждан;

14.3.5. расширять пространство для демократического представительства за счет децентрализации информационных систем и коммуникационных платформ;

14.3.6. укреплять информационную автономию граждан, улучшать способы сбора информации о политических процессах и помогать им участвовать в этих процессах удаленно, облегчая политическое выражение и обеспечивая каналы обратной связи с политическими деятелями;

14.3.7. повышать прозрачность общественной жизни и способствовать укреплению доверия между государством и обществом, а также между самими гражданами;

14.4. содержат положения о предотвращении и / или ограничении возможностей неправомерного использования ИИ для нанесения ущерба и подрыва демократии, в том числе, в частности, посредством:

14.4.1. вмешательство в избирательные процессы, персонализированное политическое преследование, формирование политического поведения избирателей и манипулирование общественным мнением;

14.4.2. усиление распространения дезинформации, «эхо-камер» и пропаганды;

14.4.3. подрыв индивидуального и общественного критического мышления;

14.4.4. содействие росту популизма и поляризации демократических обществ;

14.5. содержат положения, ограничивающие риски использования технологий на основе ИИ государствами и частными субъектами для контроля над людьми, что может привести к подрыву психологической неприкосновенности, гражданских прав и политических свобод граждан;

14.6. содержат меры предосторожности для предотвращения угрозы демократическому порядку в результате концентрации данных, информации, власти и влияния в руках нескольких крупных частных субъектов, участвующих в разработке и предоставлении технологий и услуг на основе ИИ, а также растущей зависимости отдельных лиц, учреждений и общество в целом об этих услугах и положения о том, что деятельность таких субъектов подлежит демократическому надзору.

15. Кроме того, Ассамблея считает, что для обеспечения подотчетности создаваемая правовая база должна предусматривать независимый и проактивный механизм надзора с участием всех соответствующих заинтересованных сторон, который гарантировал бы эффективное соблюдение ее положений. Такой механизм потребует наличия высококвалифицированного органа (в том числе в техническом, юридическом и этическом плане), способного отслеживать новые разработки цифровых технологий и точно и авторитетно оценивать их риски и последствия.

16. Что касается алгоритмов и платформ социальных сетей, Ассамблея считает необходимым:

16.1. сделать более прозрачными факторы принятия решений, лежащие в основе алгоритмически генерируемого контента;

16.2. даёт пользователям больше гибкости в принятии решения о том, как алгоритмы влияют на их работу в сети;

16.3. призвать платформы проводить более систематические проверки соблюдения прав человека, чтобы понять социальное влияние их алгоритмов;

16.4. рассмотреть возможность создания независимого экспертного органа для надзора за техническими платформами и работой их алгоритмов;

16.5. ужесточить контроль конфиденциальности пользовательских данных, чтобы алгоритмы имели меньше возможностей использовать данные в первую очередь.

## **В. Рекомендация проекта Примечание**

1. Парламентская ассамблея ссылается на свою резолюцию ... (2020) «Необходимость демократического управления искусственным интеллектом» и приветствует продолжающиеся усилия, предпринимаемые до сих пор на национальном, европейском и международном уровнях, в том числе на уровне Совета Европы, с целью создать нормативную базу для искусственного интеллекта (ИИ). Ассамблея отмечает, что в отсутствие такой основы частные компании, разрабатывающие и использующие технологии на основе искусственного интеллекта, до сих пор выбирали политику саморегулирования с помощью инструментов мягкого права в этой области.

2 Ассамблея приветствует работу, проделанную Специальным комитетом по искусственному интеллекту (САНАИ), которому поручено изучить осуществимость и потенциальные элементы правовой основы для разработки, разработки и применения ИИ на основе широких консультаций с участием многих заинтересованных сторон, основанный на стандартах Совета Европы в области прав человека, демократии и верховенства закона.

3 Ассамблея убеждена, что со стороны Совета Европы необходимы решительные и быстрые действия для решения проблем, возникающих в результате использования технологий на основе искусственного интеллекта, которые могут мешать, среди прочего, функционированию демократических институтов и процессов. Инструменты «мягкого права» и саморегулирования пока оказались недостаточными для решения этих проблем и защиты прав человека, демократии и верховенства закона. Совет Европы находится в стратегическом положении, чтобы обеспечить необходимое руководство и поддержку в тесном сотрудничестве и координации с другими европейскими и международными учреждениями и организациями для создания глобальной нормативной базы для ИИ.

4. В свете вышеизложенного Ассамблея рекомендует Комитету министров

4.1. выражает поддержку разработке имеющего обязательную юридическую силу документа, регулирующего искусственный интеллект, возможно, в форме конвенции, принимая во внимание, среди прочего, элементы, содержащиеся в Резолюции ... (2020);

4.2. гарантирует, что такой юридически обязательный инструмент основан на комплексном подходе, касается всего жизненного цикла систем на основе ИИ, адресован всем заинтересованным сторонам и включает механизмы, обеспечивающие реализацию этого инструмента.

## **С. Пояснительный меморандум г-жи Бергамини, докладчика**

### **Введение**

### **Порядок действий**

1. 9 апреля 2019 года Комитет по политическим вопросам и демократии инициировал предложение о принятии резолюции о «Необходимости демократического управления искусственным интеллектом». Отмечая растущее согласие в отношении того, что искусственный интеллект (ИИ) будет определяющим фактором в будущем человечества, движение подчеркивает, как ИИ уже влияет на функционирование демократии (например, вмешательство в избирательные процессы, персонализированное политическое преследование, формирование поведения избирателей и распространение дезинформации с целью манипулирования общественным мнением). В заявлении также отмечается, что концентрация власти в руках нескольких крупных частных субъектов, выходящая за рамки демократического контроля, вызывает озабоченность. Таким образом, он призывает к созданию национальной и международной нормативно-правовой базы для обеспечения демократического управления ИИ и предотвращения его неправомерного использования. Предложение было передано в наш комитет для отчета 12 апреля 2019 года, и я был назначен докладчиком 25 июня 2019 года.

2. 2 октября 2019 года комитет провел слушания с участием г-жи Биргит Шипперс, старшего преподавателя политики Университетского колледжа Святой Марии, Белфаст; Г-н Пол Немитц, главный советник, Главное управление юстиции и потребителей, Европейская комиссия; и г-н Янник Менесер, Управление информационного общества и действий против преступности, Главное управление по правам человека и соблюдению законности, Совет Европы. 27 января 2020 года

Комитет провел обмен мнениями с г-ном Дарио Фумагалли, экспертом по правовым вопросам в области защиты конфиденциальности.

3. В качестве докладчика у меня была возможность представлять Парламентскую ассамблею на мероприятиях, связанных с ИИ, включая Глобальную парламентскую сеть ОЭСР (Париж, 11 октября 2019 г.), где я выступал на сессии «Как страны подходят к своим стратегиям ИИ и политики». В мае и июне 2020 года я провел обмен мнениями (по видеоконференции) с г-ном Стивеном Фельдстайном, научным сотрудником-нерезидентом, Демократия, конфликты и управление, Фонд Карнеги за международный мир; а также три представителя европейской штаб-квартиры Facebook в Брюсселе: г-жа Мариса Хименес Мартин, заместитель директора по делам ЕС и общественной политике; Г-н Янне Эльвелид, руководитель отдела политики по делам ЕС; и г-жа Микела Палладино, специалист по государственной политике ЕС. Эти встречи были сосредоточены на этическом использовании ИИ, роли алгоритмов в платформах социальных сетей и их возможных последствиях для демократии, необходимости в нормативно-правовой базе, которая установила бы систему, основанную на ценностях и принципах, а также на расширении сотрудничества между частные компании и международные организации.

## **Обоснование отчета**

4. Являясь ключевым двигателем четвертой промышленной революции, ИИ проявляется повсюду и во всех аспектах нашей жизни. В своей воплощённой форме роботов он скоро будет управлять автомобилями, снабжать склады и заботиться о молодых и пожилых. Алгоритмы прогнозирования, присущие ИИ, окружают нас, будь то функция автоматического воспроизведения на YouTube, рекомендация фильмов на Netflix или реклама в поиске Google. Эти алгоритмы часто используются для принятия решений о ссуде, приёме в университеты и приёме на работу, но также и для работы в полиции, в аэропортах, на границах или при принятии судебных решений. Примечание. Поскольку все наши общества борются с продолжающейся пандемией Covid-19, ИИ также используется для улучшения фармацевтические исследования и помощь в анализе медицинских данных.

5. ИИ обещает решить некоторые из самых насущных проблем общества, но также создаёт проблемы, такие как непостижимые алгоритмы «чёрного ящика». Обратите внимание на потенциальную предвзятость и дискриминацию при моделировании и результатах инструментов, основанных на анализе данных, закрепление предубеждений за счёт исторических данных, потенциальное смещение работы, квази-отсутствие женщин в технических профессиях и неэтичное использование данных. Обеспечив беспрецедентный доступ к информации и обмен

ею, ИИ также усилил некоторые негативные тенденции, способствующие росту популизма и поляризации демократических обществ.

6. За последние пару лет правительства, гражданское общество, международные институты и компании вели обширные переговоры с целью определения набора общепринятых принципов того, как реагировать на сложные проблемы, создаваемые ИИ. Совет Европы, как ведущая правозащитная организация, активно участвовал в этих обсуждениях будущего ИИ и управления им. В своём последнем докладе в качестве генерального секретаря Турбьёрн Ягланд призвал к стратегическому, всестороннему подходу к ИИ, разработанному и применяемому в соответствии с европейскими стандартами в области прав человека, демократии и верховенства закона. Примечание. Комитет министров несколько раз подчёркивал, что потенциальная необходимость создания нормативно-правовой базы для ИИ. Примечание. Аналогичным образом Ассамблея указала в ряде документов на важность принятия целостного подхода и рассмотрения проблем и возможностей, связанных с ИИ в разнообразии. Примечание. В том же духе Ассамблея инициировала ряд отчётов, касающихся воздействия ИИ в различных областях.

7. В этом отчёте основное внимание уделяется тому, как ИИ влияет и влияет на функционирование демократии, а также то, как многочисленные заинтересованные стороны могут участвовать и вносить свой вклад в диалог по ИИ. Прежде всего, это аргумент в пользу создания общей основы, на которой учреждения и частные компании могут наладить открытое, ясное и готовое сотрудничество для построения общей демократической структуры управления ИИ.

## **Искусственный интеллект, определение и этические принципы**

8. Обсуждения ИИ вызвали определённое беспокойство у тех, кто опасается, что он превратится из блага для человечества в его власть. Однако не все используют одно и то же определение термина, и хотя основные элементы, как правило, одинаковы, фокус ИИ смещается в зависимости от сущности, которая даёт определение. С другой стороны, этические проблемы, связанные с ИИ, становятся все более популярными и привлекают внимание общественности по мере того, как интеллектуальные машины становятся повсеместными. Например, ИИ может моделировать и моделирует аспекты, важные для моральной воли, а именно способность отдельных лиц или коллективных образований выносить моральные суждения, принимать этические решения, основанные на представлениях о добре или зле, и нести ответственность за эти действия. Поэтому важно осознавать опасность использования ИИ для замены человеческого интеллекта в процессах принятия решений. Фактически, машинное обучение генерирует алгоритмы,

которые кажутся очень хорошими для прогнозирования, но не для понимания того, почему что-то происходит. Отсутствие причинного мышления – одна из главных проблем, которая должна побудить нас связывать ИИ строгими этическими принципами.

9. Несколько международных организаций попытались дать определение этой концепции. Согласно глоссарию Совета Европы, ИИ – это «набор наук, теорий и методов, цель которых – воспроизвести с помощью машины когнитивные способности человека». *Примечание. Европейская комиссия определяет его как «системы, демонстрирующие разумное поведение с помощью анализировать окружающую среду и предпринимать действия – с некоторой степенью автономии – для достижения конкретных целей».* *Примечание. Что касается Организации экономического сотрудничества и развития (ОЭСР), «система ИИ – это система, основанная на машинах, которая может учитывая набор целей, определённых человеком, делать прогнозы, рекомендации или решения, влияющие на реальную или виртуальную среду. Системы ИИ предназначены для работы с различными уровнями автономии».* Наконец, ЮНЕСКО отмечает, что «хотя определения ИИ не существует, дискуссии имеет тенденцию фокусироваться на машинах, способных имитировать определённые функции человеческого интеллекта, включая такие функции, как восприятие, обучение, рассуждение, решение проблем, языковое взаимодействие и даже создание креативных решений. активная работа».

10. Точно так же, хотя существует очевидное согласие с тем, что ИИ должен быть этичным, продолжаются споры о том, что представляет собой этический ИИ. За последние пять лет частные компании, исследовательские институты и организации государственного сектора издали множество наборов принципов и руководящих указаний по этическому ИИ. *Примечание. Эти руководящие принципы, как правило, согласовывают некоторые общие принципы, такие как прозрачность, справедливость, ответственность и конфиденциальность, но, похоже, резко расходятся во мнениях относительно того, что следует делать на практике. Кроме того, другие принципы, такие как устойчивость, достоинство и солидарность, значительно недопредставлены, что позволяет предположить, что эти вопросы в настоящее время остаются вне поля зрения основных этических дебатов по ИИ.*

11. Для целей всех отчётов Ассамблеи по ИИ следует понимать концепцию ИИ и этические принципы, которые должны применяться к системам ИИ, как описано в Приложении I к настоящему документу.

## **Влияние искусственного интеллекта на демократию**

12. В этом разделе, в значительной степени основанном на исследованиях Кевина Кёрнера (Deutsche Bank Research, цитируется в сноске 5) и Кателийне Мюллер, Примечание, а также на материалах, сделанных экспертами и членами комитета, я хотел бы представить карту различных способов, в которых использование технологий на основе искусственного интеллекта может и уже влияет на функционирование демократических институтов и процессов, а также на социальное и политическое поведение граждан.

13. Демократия – это управление людьми людьми ради людей. Он обеспечивает проверки против концентрации власти в руках немногих и может функционировать должным образом только в том случае, если основан на надёжных институтах, пользующихся доверием активных, преданных своему делу и информированных граждан и способных обеспечить динамический баланс интересов избирателей. Кризис современных демократий затрагивает почти все элементы демократического порядка, включая эрозию и потерю доверия к институтам, дезинформацию и дезинформацию общества, разрушение сплочённости и поляризацию общества. Современные технологии, в том числе системы на основе ИИ, могут как помочь разрешить, так и усугубить этот кризис.

14. Использование ИИ людьми не является нейтральным. Его можно использовать для усиления подотчётности правительства и может принести много преимуществ для демократических действий, участия и плюрализма, делая демократию более прямой и гибкой. Однако его также можно использовать для усиления репрессивных способностей и в целях манипулирования. Действительно, быстрая интеграция технологий искусственного интеллекта в современные средства коммуникации и платформы социальных сетей предоставляет уникальные возможности для целевого, персонализированного и часто незамеченного влияния на отдельных лиц и социальные группы, которое различные политические субъекты могут соблазнить использовать в своих интересах.

15. Опыт последних нескольких лет помогает определить некоторые ключевые области, в которых использование технологий на основе ИИ может угрожать подрывом и дестабилизацией демократии, в том числе, в частности:

- доступ к информации (дезинформация, «эхо-камеры» и размывание критического мышления);
- целенаправленное манипулирование гражданами;
- вмешательство в избирательные процессы;
- размывание гражданских прав;
- смещение финансовой и политической власти в экономике данных.

16. Более того, широкое использование государствами технологий на основе ИИ для контроля над гражданами, таких как автоматическая фильтрация информации, составляющая цензуру, и массовое наблюдение с использованием смартфонов и систем охранного телевидения в сочетании с обширными интегрированными базами данных, может привести к эрозии политических свобод, и появление цифрового авторитаризма – нового социального порядка, конкурирующего с демократией.

### **Доступ к информации – дезинформация, «эхо-камеры» и эрозия критического мышления**

17. Хорошо функционирующая демократия требует хорошо информированных граждан и предполагает, что люди с разными взглядами собираются вместе для обсуждения, чтобы найти общие решения посредством диалога. Определяя, какая информация отображается и потребляется (алгоритм веб-сайта может выборочно угадывать, какую информацию пользователь хотел бы видеть на основе информации о пользователе), технологии на основе ИИ, используемые в онлайн-СМИ, могут способствовать распространению дезинформации и разжигания ненависти, создавать «эхо-камеры» «Заметки и» пузыри фильтра «, которые приводят людей в состояние интеллектуальной изоляции, где нет места диалогу, тем самым подрывая критическое мышление и разрушая демократию. Кроме того, отдавая приоритет новостям и информации, которые нравятся пользователям, алгоритмы, как правило, укрепляют их мнения, вкусы и привычки и ограничивают доступ к разным взглядам, тем самым ограничивая свободу выбора пользователей.

18. Когда дело доходит до роли алгоритмов в распространении дезинформации и разжигания ненависти, основное внимание уделяется модерации контента, а именно тому, в какой степени алгоритмы способны выявлять и подавлять сообщения, которые нарушают стандарты сообщества и пересекают черту, когда дело доходит до распространения плохой / ложной информации. Однако не менее важным и более проблемным использованием алгоритмов такими компаниями, как Facebook, являются алгоритмы «формирования контента». Фактически, несколько платформ на основе ИИ осуществляют автоматическую цензуру (с помощью алгоритмов, определяемых владельцами) контента, публикуемого в социальных сетях частными лицами, политическими деятелями и даже государственными учреждениями, и отрицают или удаляют информацию и мнения, которые владельцы платформы не любят, тем самым ограничивая свободу слова.

19. Фактически, алгоритмы формирования контента определяют, что отдельные пользователи «видят в Интернете», включая созданные пользователями или обычные сообщения и платную рекламу. Некоторые из наиболее заметных примеров алгоритмов формирования контента включают новостную ленту Facebook, хронологию Twitter и механизм рекомендаций YouTube. Алгоритмы также определяют, каким пользователям следует показывать данную рекламу. Рекламодатель обычно устанавливает параметры таргетинга (такие как демографические данные и предполагаемые интересы), но алгоритмические системы платформы выбирают конкретных лиц, которые будут видеть рекламу, и определяют ее размещение на платформе.

20. Внутреннее исследование Facebook подтверждает эту точку зрения. Их команда пришла к выводу, что «64% всех присоединений к экстремистским группам связаны с [их] инструментами рекомендаций» и что большинство пришло из алгоритмов Facebook «Группы, к которым вы должны присоединиться» и «Откройте для себя», тем самым признав, что их «системы рекомендаций усугубляют проблему. «Примечание. Другими словами, несмотря на то что многие технологические платформы утверждают, что они проводят политику невмешательства в отношении контента, просто позволяя пользователям говорить то, что они хотят, и не вмешиваясь в их права на свободу слова, на самом деле они молча кладут руки по шкале, чтобы определить, какие сообщения будут просматривать и читать миллионы, а именно какие сообщения станут вирусными. Таким образом, их алгоритмы во многом определяют то, что пользователи видят и на что реагируют. В настоящее время, учитывая, что основным стимулом, которому следуют Facebook и другие платформы, являются доходы и прибыль, можно предположить, что даже когда контент распространяет дезинформацию, алгоритм будет повышать его видимость, если он увеличивает вовлеченность пользователей на сайт.

### **Целенаправленное манипулирование гражданами и вмешательство в избирательные процессы**

21. Хотя пропаганда и манипулирование информацией не являются новыми инструментами в политическом арсенале, коммуникационные технологии на основе ИИ значительно расширили свои масштабы и охват. Благодаря технологии на основе искусственного интеллекта онлайн и социальные сети играют все более важную роль в политическом процессе, чтобы влиять на людей и поддерживать или отвергать/отрицать партийные интересы. Некоторые тенденции, о которых сообщают политические эксперты, включают крупномасштабную скоординированную дезинформацию, в том числе с помощью «глубоких фейков»;

микротаргетирование избирателей; поляризацию общественных дебатов; подрыв доверия к демократическим институтам, политическим партиям и политикам, а также доверие общества к достоверности информации; контроль информационного потока и общественного мнения.

22. Во время выборов ИИ можно эффективно использовать для привлечения избирателей на индивидуальном уровне на протяжении всего избирательного процесса. Чат-боты и дискуссионные форумы на платформах социальных сетей, побуждающие людей оставлять комментарии/отзывы/битвы в конце, – все это различные способы измерения общественного настроения. Более того, ИИ может помочь собрать все эти данные в режиме реального времени и позволить партийным агентам изменять свои кампании, соответственно, в зависимости от того, что к ним думает общественность. Кроме того, ИИ можно использовать для манипулирования отдельными избирателями. Анализируя уникальные психографические и поведенческие профили пользователей избирателей, ИИ используется, чтобы убедить людей проголосовать за конкретного кандидата или даже создать предубеждение против оппонента этого кандидата, а также для повышения уверенности в своём выборе.

23. Хотя микротаргетинг для политических кампаний может рассматриваться просто как коммерческая реклама, он также может серьёзно угрожать демократии, общественным дебатам и выбору избирателей, когда соответствующие методы основываются на сборе данных пользователей и манипулировании ими (аналитика больших данных) для прогнозирования и влиять на их политические взгляды и результаты выборов (вычислительная пропаганда).

24. Наиболее серьёзные случаи предполагаемого вмешательства в демократические процессы на основе ИИ связаны с президентскими выборами 2016 года в Соединённых Штатах Америки. Хотя политическую консалтинговую фирму Cambridge Analytica (ныне несуществующую) обвинили в том, что она помогла Дональду Трампу победить на выборах, продвигая среди избирателей антихиллари Клинтон, некоторые крупные агрегаторы новостей и «основные средства массовой информации», как сообщалось, отдавали предпочтение новостям и видео, освещающим позитивно Клинтон и негативно изображающий Трампа.

## **Подрыв гражданских прав**

25. Доступность данных и быстрое развитие систем искусственного интеллекта повлекут за собой более широкое использование прогнозной аналитики не только компаниями, банками и рекрутерами, но также государственными учреждениями и властями. Если соответствующие недостатки и риски не будут

должным образом устранены, технологическое усиление предубеждений, а также статистических ошибок может привести к усилению исторического неравенства. Это может подорвать защиту от дискриминации и гарантии равного обращения, которые закреплены в конституциях современных демократических обществ, а также в Европейской конвенции о правах человека (ETS No. 005, статья 14) и других инструментах Совета Европы.

26. Использование систем ИИ для профилирования, отслеживания и идентификации людей, а также проверки, сортировки и даже стимулирования их поведения может оказать сдерживающее воздействие на свободу выражения мнений, свободу собраний и ассоциаций (гарантированных статьями 10 и 11 Европейской конвенции о защите прав человека). Права). Использование распознавания лиц в общественных местах может мешать свободе мнения и выражения человека просто из-за того, что защита «групповой анонимности» больше не существует. Это может отбить у людей желание посещать демонстрации и присоединяться к мирным собраниям, которые являются одним из наиболее важных элементов демократического общества. Люди также могут предпочесть воздерживаться от выражения определённых точек зрения и доступа к некоторым источникам информации, если они опасаются, что данные, собранные об их деятельности, могут быть использованы инструментами на базе искусственного интеллекта, предназначенными для принятия решений по ним (например, найм или продвижение в новое положение).

### **Концентрация власти в руках цифровых компаний**

7. Одной из наиболее общих проблем, связанных с технологиями ИИ с точки зрения демократии, является беспрецедентная и неконтролируемая концентрация данных, информации и власти в руках небольшой группы крупных цифровых компаний, которые разрабатывают алгоритмы и владеют ими, а также централизация самого интернета. Эти крупные компании больше не служат простыми каналами связи между людьми и учреждениями, а сами по себе играют все более заметную роль, определяя повестку дня, формируя и трансформируя социальные и политические модели. Если слишком много политической власти сосредоточено в нескольких частных руках, которые ставят акционерную стоимость выше общего блага, это может угрожать авторитету демократических государств. Таким образом, существует очевидная необходимость уменьшить влияние крупных частных компаний на демократический процесс принятия решений. Более того, государственно-частное сотрудничество в области ИИ и его использование в чувствительных областях, таких как общественный порядок; безопасность и разведка; пограничный контроль, а также исследования и разработки стирают

границы между обязанностями, процессами и институтами демократических государств и интересами частных корпораций.

## **Массовая слежка и усиление авторитаризма**

28. ИИ может способствовать злоупотреблению властью со стороны государства и государственных органов: как технология двойного назначения он может быть использован для подрыва важных прав человека, которые являются неотъемлемой частью функционирования демократических государств.

Примечание. Достижения в технологиях наблюдения на основе ИИ, таких как лицевые, распознавание голоса и движения вместе с сетью камер наблюдения в общественных местах позволяют отслеживать людей в реальном мире. Эти возможности искусственного интеллекта вышли на первый план во время пандемии Covid-19 (см. Сноску 6). Как и в случае с другими технологиями, инструменты наблюдения вместе с прогнозной аналитикой могут использоваться как для повышения безопасности, безопасности или контроля дорожного движения, так и для того, чтобы позволить правительствам контролировать большие толпы и прогнозировать формирование протестов и беспорядков. Управляемые сплошные меры наблюдения угрожают нашему праву на неприкосновенность частной жизни и свободу выражения мнения.

## **ИИ и принятие политических решений**

29. За последние десятилетия мы стали свидетелями определённой деполитизации процесса принятия решений. Опрос 2019 года об отношении европейцев к технологиям показал, что четверть людей предпочли бы, чтобы политические решения принимал ИИ, а не политики, независимо от того, что решения ИИ основываются на статистической корреляции доступных данных, а не на причинно-следственной связи. Между событием и решением.

Примечание. Такое мышление, вероятно, отражает растущее недоверие граждан к правительствам и политикам и подчёркивает сомнение в западной модели представительной демократии.

30. Такой подход может вызвать у избирателей пассивность, вместо того, чтобы побуждать их ставить под сомнение причины сделанного выбора и осознавать тот факт, что такой выбор основан на интересах (в благородном смысле) или ценностях, которые не обязательно должны вызывать возражения, абсолютное или «научное», чтобы считаться действительным. Это явление часто доводится до крайности и в конечном итоге приводит к использованию на риторическом уровне

формы современного *ipse dixit*. Принятие решений с помощью машин сложно, даже невозможно для людей отследить или реконструировать. Когда необъяснимые алгоритмы чёрного ящика принимают решения, которые влияют на жизнь людей, особенно в чувствительных областях, возникает серьёзная опасность для демократических ценностей, прозрачности, подотчётности и равенства, а также для принципа демократической легитимности. Пристрастие к автоматизации, то есть принятие решений, созданных машиной, без или с ограниченным человеческим контролем, подрывает прозрачность и подотчётность. Системы искусственного интеллекта могут приводить к глубоко несправедливым и даже дискриминационным результатам, которые подрывают демократические процессы и институты и негативно влияют на людей, особенно на людей из уязвимых сообществ.

31. Приучая общество к принятию выбора не на основе критического мышления, но в соответствии с велениями власти крайне несправедливо, и, следовательно, вредно, учитывая, что это невозможно установить, но всё-таки следует рассматривать общественное мнение в качестве авторитетного источника. Технологии с использованием искусственного интеллекта могут заставить людей поверить в то, что они делают свой собственный выбор, тогда как на самом деле они просто следуют шаблонам. Таким образом, ИИ может использоваться как инструмент злоупотребления прямой демократией. В более широком смысле принятие политических решений с помощью ИИ может в конечном итоге привести к установлению формы автоматизированной демократии и лишению людей автономии в отношении политических процессов. Определение политических целей нельзя оставлять на усмотрение алгоритмов, а должно оставаться в руках людей, пользующихся демократической легитимностью и берущих на себя политическую и юридическую ответственность.

32. Подводя итог, можно сказать, что технология на основе искусственного интеллекта предоставляет инструменты для вмешательства в процедуры и процессы демократий и подрыва демократических институтов. Использование ИИ и его потенциал для злоупотреблений со стороны государств и государственных агентств, а также частных корпораций представляет реальную угрозу для институтов, процессов и норм наших демократических государств, основанных на правах человека.

Примечание. Чтобы предотвратить эту угрозу: «Нам нужна структура, которая гарантирует, что эта технология будет разработана и развёрнута с полным уважением не только наших ценностей, но и верховенства наших законов, основных прав, демократии. Принцип должен заключаться в том, что ничто не может быть законным, если оно выполняется ИИ в качестве процесса автоматизации, но это было бы незаконным, если бы его выполнял человек».

## **Текущие усилия по созданию нормативной базы для искусственного интеллекта**

33. Как упоминалось выше, национальные и международные организации пытаются ответить на вопросы, связанные с использованием ИИ. В этом разделе представлен обзор действий, предпринятых крупными международными организациями, а также некоторых национальных инициатив, направленных на создание нормативной базы для ИИ.

### **Организация Объединённых Наций**

34. UNESCO готовит первый глобальный нормативный документ по этике искусственного интеллекта в соответствии с решением Генеральной конференции ЮНЕСКО на её 40-й сессии в ноябре 2019 года. Ожидается, что этот инклюзивный и междисциплинарный процесс будет включать консультации с широким кругом заинтересованных сторон, включая научное сообщество, людей различного культурного происхождения и этических взглядов, группы меньшинств, гражданского общества, правительства и частного сектора. Первая версия проекта текста рекомендации была опубликована в Интернете и открыта для консультаций. Инклюзивность, надёжность, защита окружающей среды и конфиденциальность – это принципы, включённые в настоящую Рекомендацию.

### **Европейский Союз**

35. Как описано в стратегии Европейского союза в области искусственного интеллекта, Европейская комиссия использует трёхступенчатый подход: определение ключевых требований к надёжному искусственному интеллекту; запуск крупномасштабной пилотной фазы для получения отзывов от заинтересованных сторон; и работа над достижением международного консенсуса в отношении ИИ, ориентированного на человека. Комиссия ввела семь ключевых требований к надёжному ИИ:

1. Человеческая деятельность и надзор: системы искусственного интеллекта должны способствовать созданию равноправных обществ, поддерживая деятельность человека и основные права, а не уменьшать, ограничивать или вводить в заблуждение автономию человека;

2. Надёжность и безопасность: надёжный ИИ требует, чтобы алгоритмы были безопасными, надёжными и достаточно устойчивыми, чтобы справляться с ошибками или несоответствиями на всех этапах жизненного цикла систем ИИ;

3. Конфиденциальность и управление данными: граждане должны иметь полный контроль над своими данными, при этом данные, касающиеся их, не будут использоваться для нанесения им вреда или дискриминации;

4. Прозрачность: должна быть обеспечена прослеживаемость систем ИИ;

5. разнообразие, недискриминация и справедливость: системы ИИ должны учитывать весь спектр человеческих способностей, навыков и требований и обеспечивать доступность;

6. Социальное и экологическое благополучие: системы ИИ должны использоваться для улучшения позитивных социальных изменений и повышения устойчивости и экологической ответственности;

7. Подотчётность: необходимо создать механизмы для обеспечения ответственности и подотчётности системы ИИ и её результатов.

36. Наконец, в своём White Paper Note, представленном 19 февраля 2020 года, Европейская комиссия предусматривает основу для создания надёжного ИИ, основанного на совершенстве и доверии. В партнёрстве с частным и государственным секторами. Цель состоит в мобилизации ресурсов по всей цепочке создания стоимости и создании правильных стимулов для ускорения развёртывания ИИ, в том числе малыми и средними предприятиями. Это включает в себя работу с государствами – членами и исследовательским сообществом по привлечению и удержанию талантов. Поскольку системы ИИ могут быть сложными и нести значительные риски в определённых контекстах, укрепление доверия имеет важное значение. Чёткие правила должны касаться систем ИИ с высоким уровнем риска, не возлагая слишком большую нагрузку на менее рискованные.

## **ОЭСР**

37. В мае 2019 года ОЭСР приняла Рекомендацию по искусственному интеллекту. Примечание. В Рекомендации определены пять дополнительных основанных на ценностях принципов ответственного управления надёжным ИИ:

ИИ должен приносить пользу людям и планете, способствуя инклюзивному росту, устойчивому развитию и благополучию;

Системы искусственного интеллекта должны разрабатываться таким образом, чтобы уважать верховенство закона, права человека, демократические ценности и их

разнообразии, и они должны включать соответствующие гарантии, например, возможность вмешательства человека, когда это необходимо для обеспечения справедливого общества;

В отношении систем ИИ должна быть прозрачность и ответственное раскрытие информации, чтобы люди понимали результаты, основанные на ИИ и могли оспаривать их;

Системы искусственного интеллекта должны функционировать надёжно и безопасно на протяжении всего своего жизненного цикла, а потенциальные риски должны постоянно оцениваться и контролироваться;

Организации и отдельные лица, разрабатывающие, развёртывающие или эксплуатирующие системы ИИ, должны нести ответственность за их надлежащее функционирование в соответствии с вышеуказанными принципами.

## **Действия на национальном уровне**

38. Национальные органы власти во многих странах также разработали стратегии и политику по продвижению и регулированию ИИ.

Приведу лишь несколько примеров:

Финляндия: специальная правительственная руководящая группа опубликовала в 2017 году заметку о национальной стратегии ИИ (первая страна в Европейском союзе, которая сделала это), содержащую, в частности, раздел по этике ИИ;

В 2018 году Германия опубликовала стратегию ИИ в которой основное внимание уделяется необходимости активизировать исследования и разработки, обеспечивая при этом социальную ответственность разработки ИИ;

Российская Федерация: Госдума недавно приняла национальную стратегию развития ИИ до 2030 года. В ней, в частности, упоминаются принципы на которых основывается разработка и использование ИИ (защита прав и свобод человека, безопасность и прозрачность). Упор также делается на повышение осведомлённости общественности, а также на создание интегрированной системы регулирования общественных отношений с развитием и использованием технологий искусственного интеллекта;

Примечание: Великобритания: DealNote сектора ИИ и отчёт Палаты лордов об AI>Note (оба опубликованы в апреле 2018 г.) подчёркивают важность здравого мышления и политики в отношении этики ИИ.

## Работа Совета Европы

39. Совет Европы, как ведущая правозащитная организация, играет важную роль в продвижении ИИ, соблюдающего права человека. 13 февраля 2019 года Комитет министров принял важную Декларацию о возможностях манипулирования алгоритмическими процессами. Отметим, что министры призвали государства – члены устранить риск того, что люди не смогут формировать своё мнение и принимать решения независимо от автоматизированных систем и что они могут даже подвергаться манипуляциям из-за использования передовых цифровых технологий и методов микротаргетинга. Отмечается, что инструменты машинного обучения обладают все большей способностью не только предсказывать выбор, но и влиять на эмоции и мысли, иногда подсознательно. Комитет министров призвал государства взять на себя ответственность за устранение этой растущей угрозы путём принятия соответствующих и соразмерных законодательных мер против незаконного вмешательства, расширения прав и возможностей пользователей за счёт развития важнейших навыков цифровой грамотности.

40. 26 и 27 февраля 2019 года прошла Хельсинкская конференция на тему: «Как изменить правила игры – влияние развития искусственного интеллекта на права человека, демократию и верховенство закона», организованная Советом Европы и под председательством Финляндии в Комитете министров. На конференции подчеркнули, что «должны существовать эффективные механизмы и структуры демократического надзора за проектированием, разработкой и развёртыванием ИИ», и что «функционирующие демократические процессы требуют наличия независимо информированной общественности и поощрения открытых и инклюзивных дебатов. Также необходимо повышать осведомлённость общественности о потенциальных рисках и преимуществах ИИ, а также развивать необходимые новые компетенции и навыки. Необходимо развивать общественное доверие к информационной среде и приложениям ИИ».

41. 11 сентября 2019 года Комитет министров учредил Специальный комитет по искусственному интеллекту (САНАИ) для изучения осуществимости и потенциальных элементов на основе широких консультаций с участием многих заинтересованных сторон правовой основы для разработки, разработки и применения искусственный интеллект, основанный на стандартах Совета Европы в области прав человека, демократии и верховенства закона.

42. Рекомендация CM/Rec (2020) 1 Комитета министров государствам –членам о воздействии алгоритмических систем на права человека, принятая 8 апреля 2020 года в контексте пандемии Covid-19, выпустила набор руководящих принципов, призывающих правительства гарантировать, что они не нарушают права человека из-за собственного использования, разработки или приобретения алгоритмических

систем. Кроме того, как регулирующие органы они должны создавать эффективные и предсказуемые законодательные, регулирующие и надзорные основы, которые предотвращают, выявляют, запрещают и устраняют нарушения прав человека, исходящие от государственных или частных субъектов.

### **Искусственный интеллект и крупные компании: ответственность и этика**

43. Этические инициативы помогают выработать общий язык для обсуждения социальных и политических проблем. Они предоставляют разработчикам, сотрудникам компании и другим заинтересованным сторонам набор высокоуровневых заявлений о ценностях, по которым впоследствии можно будет оценивать действия. Они также являются образовательными, часто выполняя работу по повышению осведомлённости о конкретных рисках ИИ как внутри данного учреждения, так и за его пределами, среди более широкой заинтересованной общественности. Но они не подходят для обеспечения подотчётности.

44. Как подчёркивается в отчёте, опубликованном AI Now Institute Note в Нью-Йоркском университете в 2019 году, компании, создающие решения на основе искусственного интеллекта для всего, от оценки студентов до оценки иммигрантов на предмет преступности, связаны немногим, кроме нескольких этических заявлений, которые они приняли для себя. Хотя компании прилагали определённые усилия, авторы утверждают, что «структуры, управляющие в настоящее время ИИ, не способны обеспечить подотчётность». «По мере того как распространённость, сложность и масштаб этих систем растут, отсутствие значимой подотчётности и надзора, включая основные гарантии ответственности и надлежащей правовой процедуры, становится все более насущной проблемой».

45. В 2018 году генеральный директор Google Сундар Пичаи опубликовал общедоступный набор из семи «руководящих принципов», призванных гарантировать, что работа компании над ИИ будет социально ответственной. Эти этические принципы включают обязательство «быть социально выгодным» и «избегать создания или усиления несправедливой предвзятости». В статье для «Financial Times» г-н Пичаи призвал к регулированию ИИ, но высказался за более разумный подход, поддерживая то, что отдельные области развития ИИ, такие как беспилотные автомобили и медицинские технологии, требуют индивидуальных правил, упомянув, что Google запустил свой собственный независимый совет по этике в 2019 году, но закрыл его менее чем через две недели после разногласий по поводу того, кто в него был назначен.

46. Другие компании, в том числе Microsoft, Facebook и производитель видеочамер для полиции Axon, также собрали советы по этике, консультантов и

группы. Такие события обнадеживают, и стоит отметить, что те, кто стоит в центре разработки ИИ, заявили, что серьёзно относятся к этике. Однако, как указано выше, они не обеспечивают прочной основы для привлечения компаний к ответственности.

## **Выводы**

47. Демократия подразумевает, что люди с разными взглядами объединяются для поиска общих решений посредством диалога. Вместо того чтобы создавать общественное пространство и общую повестку дня, коммуникационные платформы на основе ИИ, похоже, отдают предпочтение индивидуалистическим и поляризованным взглядам и приводят к появлению закрытых интернет-сообществ, разделяющих те же взгляды, тем самым подрывая социальную сплочённость и демократические дискуссии и, напротив, способствуя распространению языка вражды, раздроблению и сегментации общества. Тот факт, что отдельные слои населения не используют платформы из-за различных пробелов в использовании информационных и коммуникационных технологий (ИКТ) (а именно, в зависимости от пола, возраста, социального происхождения), также необходимо учитывать при этом размышлении. В Европейском Союзе, например, гендерный разрыв в цифровых навыках составляет 11%, с более высоким разрывом в навыках, превышающих базовые, и особенно среди лиц старше 55 лет. Обратите внимание на частные компании, которые применяют правила рынка, а не правила демократии и не несут ответственности за разжигание ненависти и распространение материалов с насилием.

48. Технологии на основе ИИ вмешиваются в функционирование демократических институтов и процессов и влияют на социальное и политическое поведение граждан. Все алгоритмы машинного обучения работают на основе структуры классификатора, в которой машина учится делать набор предположений о различных цепочках данных. Как и все процессы интерактивного обучения, машинное обучение также может страдать от ложноотрицательных или ложноположительных отчётов. Хотя эти отчёты являются распространёнными ошибками в любом таком исследовании, как только такая погрешность переносится на политические решения, это может привести к систематическим репрессиям определённых этнических или социальных групп, неправомерному привлечению подозреваемых или ненужному систематическому профилированию граждан.

49. Существует очевидный разрыв между темпами технологического развития и нормативно-правовой базой. Принципы и политика саморегулирования не могут быть единственными инструментами регулирования ИИ, поскольку они не ведут к ответственности. Европе необходимо обеспечить, чтобы возможности ИИ

регулировались и использовались для общего блага. Следовательно, существует необходимость в создании нормативной базы для ИИ с конкретными принципами, основанными на защите прав человека, демократии и верховенстве закона. Любая работа в этой области должна вовлекать все заинтересованные стороны, включая, в частности, граждан и частные компании. Работа САНАИ, которая в конечном итоге должна привести к созданию правовых рамок для демократического управления искусственным интеллектом на основе стандартов Совета Европы, нуждается в полной поддержке и поощрении.

50. Для обеспечения подотчётности создаваемая правовая база должна предусматривать независимый механизм надзора, который гарантировал бы эффективное соблюдение её положений. Без такого механизма крупные компании ИКТ просто продолжали бы вести свою обычную деятельность, и, несмотря на огромную мощь и транснациональный характер этих компаний, большинство государств продолжало бы закрывать глаза на несоблюдение правил поведения как на необходимую и приемлемую цену для преследования жизненно важных интересов.

51. Однако такой механизм надзора может быть эффективным только в том случае, если он может быть активным и задействованным заранее. Действительно, хотя было бы важно ввести санкции за несоблюдение правил поведения, механизм, который ограничился бы штрафами и штрафами *ex post*, которые обычно легко доступны для крупных частных компаний независимо от суммы, не принесёт желаемого результата. Это связано с тем, что часто очень сложно, если не невозможно, восстановить предыдущую ситуацию или «стереть ущерб» после того, как данная технология искусственного интеллекта была внедрена и использована как неэтичная и/или не соответствующая правам человека, демократии и верховенству закона.

52. Для упреждающего механизма надзора требуется высококвалифицированный орган (в том числе в техническом, юридическом и этическом плане), способный отслеживать новые разработки цифровых технологий и точно и авторитетно оценивать их риски и последствия. Само собой разумеется, что в такой орган должны быть вовлечены все соответствующие заинтересованные стороны.

53. Что ещё более важно, роль ИИ в изменении баланса сил между институтами, политическими деятелями и исполнительными органами требует более структурированного исследования. Учитывая масштаб проблем легитимности и суверенитета, связанных с передачей политических решений алгоритмам, роль конституций, парламентов и политических элит по отношению к ИИ необходимо тщательно изучить с особым акцентом на том, как политическая власть должна выражаться в автоматизированных решениях.

54. Это не означает, что ИИ не может быть силой добра или делать политику более эффективной или более отзывчивой к нуждам граждан. При правильном использовании ИИ может расширить пространство для демократического представительства за счёт децентрализации информационных систем и коммуникационных платформ. Это может усилить информационную автономию граждан и улучшить способы сбора информации о политических процессах и помочь им удалённо участвовать в этих процессах. Подобно тому, как ИИ можно использовать для усиления непрозрачности и безответственности, он также может повысить прозрачность и способствовать укреплению доверия между государством и обществом, а также между самими гражданами.

55. Со своей стороны, Совет Европы как ведущая международная организация, устанавливающая стандарты в области демократии, должен играть новаторскую роль в разработке способов и форматов, обеспечивающих использование технологий на основе ИИ для укрепления демократии через собрания граждан, электронные агоры и другие формы совещательного и совместного участия людей в демократических процессах.

## **Правовые аспекты автономных транспортных средств**

Сообщить | Док. 15143 / 22 сентября 2020 года

Комитет по правовым вопросам и правам человека

Докладчик: г-н Зия АЛТУНЬЯЛДЫЗ, Турция, NR

Происхождение: Ссылка комитету: Док. 14839, ссылка 4442 от 12 апреля 2019 года. 2020.

### **Краткие сведения**

Движение полуавтономных транспортных средств по европейским дорогам вероятно, в ближайшие годы значительно увеличится, причём полностью автономные транспортные средства могут стать доступными в течение следующего десятилетия. Эти события вызывают вопросы в отношении уголовной и гражданско-правовой ответственности, обязательств производителей и страховщиков и будущего регулирования автомобильных перевозок. Возникают также вопросы этики и конфиденциальности.

Современные автоматизированные приводные системы отличаются своими опорами на системы искусственного интеллекта. Ожидается, что автоматизированные транспортные средства будут значительно безопаснее, чем те, которые управляются людьми. Для этого потребуются соответствующее регулирование и потенциал. В качестве отправной точки данный регламент должен обеспечить полное уважение права на жизнь.

Комитет по правовым вопросам и правам человека предлагает, чтобы разработка и внедрение автономных транспортных средств регулировались в соответствии со стандартами Совета Европы в области прав человека и верховенством закона. Также предлагается, чтобы Комитет министров принял во внимание особо серьёзное потенциальное воздействие на человека права на использование искусственного интеллекта в автоматизированном вождении при оценке необходимости и целесообразности создания правовой базы.

## **Введение**

### **А. Проект резолюции**

1. Движение полуавтономных транспортных средств на европейских дорогах в ближайшие годы значительно увеличится. Возможно, что полностью автономные транспортные средства могут появиться в продаже в течение следующего десятилетия. Эти события вызывают вопросы в отношении уголовной и гражданской ответственности, ответственности производителей и страховщиков, а также будущего регулирования автомобильных перевозок. Также вызывают опасения вопросы этики и конфиденциальности.

2. В случае полуавтономного транспортного средства, работающего под контролем автоматизированной системы вождения (ADS) или полностью автономного транспортного средства, уголовное право не предназначено для борьбы с поведением нечеловеческих субъектов. Это может создать «пробел в ответственности», где человек в транспортном средстве – ответственный пользователь», даже если на самом деле не участвовал в вождении – не может быть привлечён к ответственности за преступное деяние, а само транспортное средство работало в соответствии с требованиями производителей и применяемыми правилами. Это может потребовать новых подходов к распределению уголовной ответственности или альтернативам уголовной ответственности в случаях, когда ни один человек не может быть привлечён к ответственности.

3. Аналогичные опасения касаются и гражданской ответственности за ущерб, причинённый транспортным средствам, эксплуатируемым под надлежащим контролем ADS. Текущие режимы ответственности на основе вины могут освободить ответственного пользователя от какой-либо ответственности, а ответственность будет передана ADS. Это может потребовать новых подходов, таких как строгая ответственность, чтобы гарантировать, что потерпевшие стороны получают компенсацию за причинённый им ущерб.

4. В случаях, когда правила дорожного движения нарушаются транспортным средством, находящимся под надлежащим контролем ADS, будь то установление фактов совершения уголовного преступления или причинения ущерба третьим лицам, ответственность производителя может вызвать вопросы за качество продукции. Однако сложность автономных транспортных средств может затруднить доказательство существования и характера какой-либо технической неисправности. Опять же важно, чтобы будущие правила не оставляли пробелов в этом отношении.

5. Эти опасения тесно связаны с этическими проблемами, возникающими в связи с технологиями автономных транспортных средств. Водители-люди регулярно обязаны принимать этические решения, в том числе вынужденный выбор решения о жизни и смерти. Компания ADS должна будет принимать те же решения, но в соответствии с этическими принципами, установленными их производителем. Учитывая, что покупатели автономных транспортных средств могут предпочесть, чтобы приоритет был отдан их собственной безопасности, давление со стороны конкурентного рынка на производителей может не привести к результатам, оптимальным с общей утилитарной точки зрения. Несомненно, может потребоваться государственное регулирование для стандартизации этических решений, подразумеваемых при разработке ADS, для обеспечения совместимости с интересами широкой общественности.

6. ADS зависят от данных и генерируют данные, в том числе персональные данные, касающиеся, например, передвижения человека. Данные от автономных транспортных средств автоматически передаются другим автономным транспортным средствам и в центральную систему, и при определённых обстоятельствах может потребоваться обмен данными с регулируемыми и правоохранительными органами. Особое внимание потребуется для обеспечения правильного баланса между обработкой данных, необходимой для безопасной эксплуатации автономных транспортных средств, и уважением, защитой конфиденциальности водителей, пассажиров и других пользователей.

7. Современные ADS отличаются тем, что они полагаются на системы искусственного интеллекта (ИИ); действительно, современные автономные транспортные средства по сути, являются роботами. Внедрение автономных транспортных средств означает постановку управляемых ИИ роботов на зарядку в ситуации доказанного серьёзного потенциального риска для своих пассажиров и

других участников дорожного движения. Ожидается, что автономные транспортные средства будут значительно более безопасными, иметь потенциал быть значительно безопаснее, чем те, которые управляются людьми. Для этого потребуется соответствующее регулирование. В качестве отправной точки данное положение должно обеспечить полное уважение права на жизнь, включая позитивные обязательства по предотвращению предсказуемых угроз.

8. Парламентская Ассамблея считает, что этические и нормативные стандарты, применимые к ИИ в целом, также должны применяться к его использованию в автономных транспортных средствах. Также считается, что работа специального Комитета по искусственному интеллекту (САНАИ) над возможной правовой базой для ИИ будет особенно значимой, и отмечается важный вклад в работу в этой области других международных организаций, включая Организацию экономического сотрудничества, Европейский Союз и Организацию Объединённых Наций.

9. Когда станут доступны полностью автономные транспортные средства, разработанные для соблюдения правил дорожного движения и предотвращения любых столкновений, законодатель должен будет решить проблемы, возникающие в результате их сосуществования с транспортными средствами, управляемыми людьми, которые не всегда соблюдают правила. Демократический законодатель должен будет принять решение о наиболее подходящем балансе между минимизацией числа жертв несчастных случаев и обеспечением возможности для эффективного потока трафика.

10. Ассамблея приходит к выводу, что приведённые выше соображения порождают множество новых проблем для режимов регулирования. Принимается к сведению работа, проводимая в специализированных регулирующих органах, включая рабочую группу по автономным и подключённым транспортным средствам (GRVA) Европейской экономической комиссии Организации Объединённых Наций, которая решает целый ряд важнейших технических вопросов, а также Европейский Союз и различные национальные органы власти. Отмечается работа в рамках Совета Европы по теме «Искусственный интеллект и уголовно-правовая ответственность в государствах – членах Совета Европы в случае с автоматизированными транспортными средствами», ведущаяся в Европейском комитете о проблемах преступности (CDPC).

11. Ассамблея призывает:

11.1. Государства – члены Совета Европы обеспечить, чтобы последствия разработки и внедрения автономных транспортных средств для уголовного, гражданского права и прав человека регулировались в соответствии со стандартами Совета Европы в области прав человека и верховенства закона, включая уважение

права на жизнь, неприкосновенность частной жизни и принцип правовой определённости;

11.2. GRVA провести оценку воздействия на права человека в рамках своей подготовительной работы по будущему регулированию автономных транспортных средств, как часть общей, всеобъемлющей основы для обеспечения максимальной безопасности во всех её во время будущей разработки и производства автономных транспортных средств;

11.3. СВЗС обеспечить выявление и устранение возможных пробелов в применении уголовного законодательства к эксплуатации автономных транспортных средств и адресовано;

11.4. САНАИ обратить особое внимание применению ИИ в ADS, где существует особый риск неблагоприятных последствий для осуществления основных прав человека, при составлении карты рисков и возможностей ИИ.

## **В. Проект рекомендации**

1. Парламентская Ассамблея ссылается на свою резолюцию (2020) о правовых аспектах «автономных» транспортных средств. Напоминает, что эта резолюция была принята, поскольку соответствующая работа в Совете Европы продолжалась специальным комитетом по искусственный интеллект (САНАИ).

2. Ассамблея призывает Комитет министров принять во внимание особо серьёзное потенциальное воздействие на права человека использование искусственного интеллекта в автоматизированных системах вождения при оценке необходимости и применения правовой базы для искусственного интеллекта.

## **С. Пояснительная записка г-на Алтуньялдиза, докладчик**

### **Введение**

### **Предыстория**

1. Ходатайство о принятии резолюции (Дос. 14839), лежащее в основе этого доклада, который я представил 7 марта 2019 года, был передан Комитету 12 апреля 2019 года, после чего 29 мая 2019 года комитет назначил меня докладчиком.

Комитет провёл слушания с участием экспертов: г-жи Теодоры Хамсен, Федеральное министерство транспорта и цифровой инфраструктуры Германии, профессора Сахина Албайрака, исполнительного директора Лаборатории распределённого искусственного интеллекта (DAI-Lab), Берлин и г-на Коннора Шампа, группа по автоматизированным транспортным средствам Комиссии по праву, Лондон. Я хотел бы поблагодарить всех трёх экспертов за их вклад в подготовку настоящего доклада. Я намеревался провести ознакомительный визит в научно-исследовательский центр, работающий над технологиями автономных транспортных средств, но, к сожалению, ограничения из-за пандемии Covid-19 сделали это невозможным.

2. Технический прогресс уже сделал распространение полуавтономных транспортных средств реальностью. Как отмечается в ходатайстве, как увеличившийся оборот полуавтономных транспортных средств, так и ожидаемое распространение полностью автономных транспортных средств создают вопросы в отношении уголовной и гражданской ответственности, обязательств производителей и будущего регулирования перевозки автомобилей, которые необходимо будет решить законодательными органами государств – членов Совета Европы и других стран. С этими вопросами связан ряд этических проблем и проблем конфиденциальности, которые также необходимо будет решить.

3. Гонка за разработкой транспортных средств с повышением уровня автономности привело к первому смертельному исходу в мае 2016 года. Водитель, который включил режим автопилота Tesla, врезался в большой белый 18-колесный грузовик с прицепом, пересекающий шоссе, обнаружить который в ярких условиях не удалось. В ходе расследования, проведённого Национальной администрацией безопасности дорожного движения США, установлено, что компания Tesla не виновата. Первый случайный прохожий был убит Volvo XC-90 в марте 2018 года в ходе использования его компанией Uber для тестирования своей технологии автономного вождения. Автомобиль обнаружил пешехода, но его система экстренного торможения была отключена. Прокуроры установили, что Uber уголовной ответственности не несёт, хотя не исключено, что резервный водитель безопасности будет привлечён к уголовной ответственности.

4. Технологии вспомогательного вождения постоянно развиваются. Тем не менее, прежде чем полностью автономные транспортные средства станут коммерчески доступны и разрешены на дорогах общего пользования, всё ещё требуется значительный технический и нормативный прогресс. Европейский научно-исследовательский Консультативный совет по автомобильному транспорту (ERTRAC) прогнозирует, что полностью автоматизированные шаттлы и автобусы, работающие в пределах определённых городских условий, станут доступны не раньше 2030 года, а полностью автоматизированные легковые автомобили будут доступны только после 2030 года.

## **Ключ концепции-автоматизация транспортных средств**

5. Транспортные средства, оснащённые автоматизированными системами вождения, часто в просторечии называют «автономными». Однако, для того чтобы иметь полное представление об этических и юридических вопросах важно уточнить различия между различными уровнями технологий. SAE International (ранее общество Автомобильных инженеров) определили шесть уровней автоматизации вождения в Международном стандарте J3016 для потребителей, и их классификация получила широкое распространение.

6. Уровни начинаются с уровня 0, который относится к транспортным средствам, которые полностью управляются людьми и заканчиваются уровнем 5, который относится к транспортным средствам, которые полностью автономны при любых обстоятельствах, не требуя вмешательства человека. Между этими классификациями транспортных средств находятся системы помощи водителю, например гидроусилитель руля и антиблокировочная тормозная система (уровень 1); частичная автоматизация, например, системы автоматического торможения для обнаружения и предотвращения неминуемых столкновений (уровень 2); условная автоматика, при которой система дополнительно контролирует среду вождения и подсказывает вмешательство водителя (уровень 3); и высокая автоматизация, когда система способна управлять автомобилем даже при отсутствии вмешательства водителя (уровень 4). Техническое выражение «полуавтономные транспортные средства» относится к автоматизация 3-го и 4-го уровней. Мой доклад в основном касается полуавтономных транспортных средств, но в нём также рассматриваются юридические проблемы, возникающие в связи с полностью автономными транспортными средствами (уровень 5).

7. Текущий максимальный общедоступный уровень автоматизации транспортных средств – это уровень 3. Эти автомобили имеют возможность обнаружения окружающей среды, но все ещё требуют вмешательства человека. Водитель должен сохранять бдительность и брать на себя управление, если система не в состоянии выполнить задачу. Технология уровня 4 существует в качестве прототипа и в настоящее время находится в стадии тестирования; некоторые производители утверждают, что их самые последние модели уже включают в себя технология уровня 4, хотя это, по-видимому, основано на щедром толкование данного определения и не принимается независимыми экспертами. Эти транспортные средства способны вмешаться в случае сбоя системы. В большинстве случаев не требуют человеческого вмешательства. Однако, водитель все ещё имеет возможность вручную управлять автомобилем. Этот тип технологии используются в настоящее время, но только по заранее предопределённым маршрутам и при определённых обстоятельствах. Например, автобусы-шатлы, использующие технологию уровня 4, используются в некоторых пенсионных учреждениях

сообщества, университетских городах и аэропортах. Полная автоматизация 5 уровня в настоящее время невозможна.

### **Автономные транспортные средства и искусственный интеллект**

8. Современные автономные транспортные средств полностью зависят от систем искусственного интеллекта (ИИ). Усовершенствованные сенсорные системы, в частности LiDAR, (обнаружение света и дальность действия) и радар, предоставляют подробную информацию на 360° о рабочей среде автомобиля. Эта информация, наряду с информацией со спутниковых систем позиционирования и бортовых цифровых карт должна быть обработана таким образом, чтобы транспортное средство могло определить своё местоположение, планировать и следовать по маршруту, распознавать и надлежащим образом реагировать на разметку, такую как дорожные знаки и линии и опасности, как реальные и даже потенциальные препятствия на пути движения транспортного средства, включая других участников дорожного движения (моторизованные транспортные средства, велосипедисты и пешеходы). Эта обработка выполняется с помощью алгоритмов машинного обучения ИИ, обученных на огромных исторических наборах данных и постоянно совершенствующих свою собственную производительность с помощью накопленного опыта (включая опыт эксплуатации других транспортных средств, работающих с той же системой). Для автоматизации уровня 3 и выше система искусственного интеллекта будет полностью под контролем автомобиля по крайней мере некоторое время, и решения, которые она принимает, могут быть буквально вопросом жизни и смерти. Как отмечалось выше, даже на уровне 3 автоматизация и во время тестирования систем уровня 4 с участием человека, несчастные случаи со смертельным исходом имели место.

9. Профессор Албайрак выделил важное различие в способах разработки будущих систем автономных транспортных средств. Самые известные из них проекты, например, Google Waymo или Tesla предполагают «умные» транспортные средства, движущиеся по «тупым» дорогам, и таким образом полностью зависят от их собственных возможностей восприятия и обработки данных. Профессор Албайрак разрабатывает другой подход, используя технологии, встроенные не только в автомобиль, но и в дорогу и её инфраструктуру. Статические камеры и другие датчики следят за самой дорогой и движением по ней. Исследовательская группа профессора Албайрака в настоящее время экспериментирует с элементами такой системы на реальном полигоне в центре Берлина. Преимущества такой интегрированной системы с точки зрения общего управления трафиком кажется очевидной, особенно в условиях плотной городской среды, хотя её постоянные затраты на инфраструктуру и относительная полезность могут сделать её менее

актуальной на сельских дорогах. Тем не менее, как указал, Мистер Чэмп есть также законные сомнения в пригодности систем уровня 4/5 для движения по небольшим сельским дорогам, где экологические условия весьма различны и сложны по-разному чем в городах.

10. Автоматизированные транспортные средства – это, по сути, роботы, а по существу, автономная машина с компьютерным управлением, предназначенная для выполнения определённой функции автономно. Возложение на роботов ответственности за перевозку пассажиров-людей на дорогах общего пользования имеет огромное значение для безопасности.

В 2018 году на дорогах государств-членов ЕС было 268 миллионов легковых автомобилей, более 33 миллионов фургонов и 6,6 миллионов грузовиков. На тех же дорогах погибло 25 000 человек. Внедрение автоматизированных транспортных средств означает передачу ИИ контроля за быстроходными пассажирскими снарядами в условиях серьёзной опасности как для пассажиров, так и для других участников дорожного движения. Очевидно, ожидается, что автоматизированные транспортные средства окажутся безопаснее, чем те, которые управляются людьми. Однако, прежде чем это будет обеспечено, необходимо принять серьёзные меры регулирования и предосторожности. Таким образом, с точки зрения прав человека, ключевым соображением будет обеспечение того, что автоматизированные транспортные средства и системы искусственного интеллекта, которые ими управляют, будут регулироваться таким образом, чтобы обеспечить полное уважение права на жизнь, включая позитивные обязательства по предотвращению предсказуемых угроз.

11. Совет Европы уже приступил к работе над вопросами уголовного права, касающимися применения ИИ в контексте систем автономных транспортных средств. Я рассмотрю эту работу, которая проводится в межправительственном Европейском Комитете по проблемам преступности (CDPC), более подробно об этом ниже.

12. Параллельно с работой CDPC в сентябре 2019 года Комитет Министров учредил межправительственный специальный комитет по искусственному интеллекту (СНАИ).

СНАИ было поручено изучить целесообразность и потенциальные элементы правовой базы для проектирования, разработки и применения искусственного интеллекта. Его работа основана на стандартах Совета Европы в области демократии, прав человека и верховенства, а также на других соответствующих международных документах и на текущей работе других международных и региональных организаций. Наряду с обычными участниками, представляющими государства – члены Совета Европы, государства – наблюдатели и другие органы Совета Европы органы (включая Ассамблею), СНАИ имеет исключительно

высокий уровень участия представителей органов частного сектора, гражданского общества, научно-исследовательских и академических институтов.

13. САНАИ провёл своё первое заседание 18-20 ноября 2019 года. Среди прочего было решено, что ключевым элементом будущего технико-экономического обоснования будет «отображение рисков и возможностей, возникающих в результате разработки, проектирования и применения искусственного интеллекта, включая влияние последнего на права человека, верховенство права и демократию». Принятие технико-экономического обоснования ожидается на третьем совещании САНАИ, запланированном на декабрь 2020 года.

14. Это институциональный контекст, в рамках которого Ассамблея будет обсуждать настоящий и другие доклады, связанные с искусственным интеллектом, которые в настоящее время готовятся различными комитетами. Ассамблея решила подойти к теме на контекстуальной основе, изучая влияние искусственного интеллекта в различных областях. В Комитете по правовым вопросам и правам человека, например, есть также отчёты о воздействии ИИ на полицейскую деятельность и систему уголовного правосудия, на технологию интерфейса мозга-компьютера (на ранних стадиях подготовки) на смертоносные автономные системы вооружения. Рекомендации, которые Ассамблея может принять на основе этих докладов будет таким образом обеспечивать важное руководство для САНАИ при котором составление карт рисков и возможностей ИИ и его влияние на права человека будут соответствовать верховенству закона и демократии.

15. Следует также отметить, что другие международные организации также работают над этическим и/или правовым регулированием ИИ с различной направленностью и подходом в зависимости от их институциональной перспективы. В Европе, например, Европейский Союз разработал Европейскую стратегию по искусственному интеллекту, реализация которой поддерживается экспертной группой высокого уровня по искусственному интеллекту, а ОЭСР приняла принципы по искусственному интеллекту с сильной этической составляющей.

16. Далее, общая информация об ИИ, включая описание и изучение применимых этических принципов может быть найдено в приложении к настоящему докладу.

## **Этические проблемы**

17. Введение и возможное распространение полуавтономных и автономных транспортных средств поднимают ряд этических вопросов. Вождение – это не просто следование правилам дорожного движения. Водители регулярно обязаны делать то, что является, по сути, этическими решениями, особенно в ситуациях

вынужденного выбора, таких как некоторые неизбежные столкновения. По мере повышения уровня автоматизации возрастает так же и способность принимать решения. Машины будут запрограммированы с помощью сложных алгоритмов принудительного выбора, позволяющих им принимать этические решения, например, о том, лучше ли сбить двух пешеходов или двух велосипедистов. Этические допущения, присущие таким алгоритмам, могут быть проблематичным. В настоящее время ведутся споры вокруг основы критериев принятия решений в этих алгоритмах и о том, должны ли правительства регулировать эти критерии для стандартизации или устанавливать минимальные моральные стандарты. В случае неизбежного – аварии могут иметь значение несколько факторов, такие как количество пострадавших людей, степень тяжести и вероятность различных видов травм и, возможно, личностные характеристики, такие как возраст и инвалидность.

18. Ключевая этическая проблема в связи с развитием автономных систем заключается в том, что с коммерческой точки зрения производителям может быть более выгодно проектировать транспортные средства, в которых приоритетом является безопасность автомобиля и его пассажиров, так как люди могут предпочесть покупать автомобили, которые будут держать их в безопасности. Однако с утилитарной точки зрения, это не всегда приводит к наиболее этичному образу действий. Необходимо найти баланс между свободой производителей внедрять инновации и желанием получить коммерческую прибыль с общественными соображениями общественной безопасности.

19. С социально-экономической точки зрения существуют опасения по поводу доступности технологии автономного вождения. Ключевым преимуществом разработки автономных транспортных средств будет сокращение дорожно-транспортных происшествий, поскольку более 90% дорожно-транспортных происшествий вызваны человеческими ошибками. Это в свою очередь должно привести к снижению страховых взносов для владельцев полуавтономных и автономных транспортных средств. Однако, эта новая технология в настоящее время доступна только для самых богатых членов общества, а это означает, что люди из наиболее низких социально-экономических классов с большей вероятностью будут управлять автомобилями, которые менее безопасны и увидят, что их страховые взносы растут, по крайней мере, относительно.

20. Несмотря на то, что существует широкое общее понимание, в том числе на международном уровне, актуальности и значения основных этических принципов, применимых к автономным транспортным средствам, необходимо обеспечить, чтобы это понимание было переведено в правила, которые в равной степени широко принимаются и выполняются.

## Вопросы ответственности

21. В случае, если ущерб будет причинён полуавтономным или автономным транспортным средством, неизбежно возникнут вопросы о том, кто может быть привлечён к уголовной или гражданской ответственности, и при каких обстоятельствах. Ответ на эти вопросы неизменно будет зависеть от уровня автоматизации вождения. В отношении транспортных средств уровня 3 или 4 проблемы с ответственностью особенно остро стоят во время переходных этапов между автоматическим и ручным режимами вождения. В случае транспортных средств уровня 5 возможен спор о том, будет ли принято окончательное решение о передаче управления водителю-человеку или машине.

## Уголовная ответственность

22. CDPC определила сложность борьбы с «преступным поведением» нечеловеческих существ в качестве ключевой проблемы в современном уголовном праве. Если транспортным средством управляет машина, а не человек, существует риск возникновения «разрыва в ответственности». Понятие вины или *mens rea* имеет решающее значение в европейских системах уголовного права. Там, где авария связана с полуавтономным транспортным средством, распределение вины между водителем и системой может быть чрезвычайно сложной задачей. В настоящее время в большинстве европейских государств если используется система помощи при вождении и происходит авария, водитель все равно может быть обвинён в халатности, поскольку он обязан следить за системой. При отсутствии халатности, производитель может быть привлечён к уголовной ответственности. Для автоматизации уровня 3 и 4 точное разграничение между тем когда человек-водитель считается ответственным (и, следовательно, потенциально несёт ответственность согласно уголовному законодательству), и когда нет, необходимо будет чётко определить в законе на основе принципов, установленных на международном уровне.

23. В будущем, если технология уровня 5 станет коммерчески доступной, необходимо будет разработать новые способы установления вины (уголовной ответственности) в уголовно-процессуальном праве, возможно, вместе с установлением новых составов преступлений в материальном уголовном праве. Этот может вызвать вопрос о том, является ли нечеловеческая сущность, а именно ИИ, ответственным за управление транспортным средством, аналогично тому, как корпорация может нести ответственность как «юридическое лицо». Ясно, что это будет иметь очень сложные концептуальные и правовые последствия.

24. Кроме того, не все вопросы, касающиеся уголовного права и технологии автономного вождения, относятся к авариям, вызванным этими транспортными средствами. Снижение ответственности водителя также может повлиять на применимость других правонарушений, таких как вождение в нетрезвом виде или использование мобильных телефонов во время вождения. В этом отношении различные уровни автоматизации движения затрудняют разработку комплексной системы регулирования.

25. Юридическая комиссия Соединённого Королевства в настоящее время проводит детальную исследовательскую работу по правовым проблемам, связанным с внедрением автономных транспортных средств, включая уголовно-правовые последствия. В ходе серии публичных консультаций Юридическая комиссия хотела получить комментарии по широкому кругу вопросов и предложений по будущему регулированию автономных транспортных средств. Эта работа раскрывает широкий спектр новых и иногда удивительных проблем, которые возникают в отношении уголовно-правового регулирования автономных транспортных средств.

Например: для автоматизации уровней 3 и 4, там, где всё ещё есть человек, «ответственный за» (если использовать выражение Юридической комиссии), проблема отвлечения внимания водителя и, как следствие, отсутствие ситуационной осведомлённости и запоздалой реакции, были предсказуемой проблемой, но большинство считало, что не должно быть «никакого ослабления законов против отвлечённого вождения». Производители, однако, «поддержали разрешение некоторых видов деятельности, кроме вождения автомобиля» для ответственного пользователя.

В случае аварии, произошедшей при управлении системы автономного транспортного средства, страховщикам понадобится доступ к данным о событии. В связи с отсутствием автоматического сообщения об авариях (самой автономной системой транспортного средства) возникает вопрос о том, обязан ли ответственный пользователь или застрахованный владелец уведомлять об аварии и в какие сроки.

В тех случаях, когда система автономного транспортного средства находилась в «ведении» во время нарушения правил вождения, Юридическая комиссия предложила, чтобы полиция передала дело в регулирующий орган, с правом применять регулирующие санкции к «автоматизированному вождению». Таким образом можно было бы обойти вопрос о том, может ли ИИ подпадать под действие норм уголовного права, которые обычно закреплены за людьми, и которые зависят от понятий личной автономии и моральной ответственности.

Вопросы о допустимых исключениях из нормальных правил вождения разделились во мнениях: следует ли когда-нибудь разрешать автономному транспортному средству подъезжать к тротуару (например, разрешить проезжать машинам скорой помощи); следует ли когда-либо разрешать превышение

скоростного режима (Ассоциации водителей считали, что это может быть позволено, до определённого момента); и может ли он когда-нибудь медленно продвигаться через толпу пешеходов (большинство комментариев гласило, что этого почти никогда нельзя допускать, что в противном случае автономные транспортные средства могут быть слишком легко заблокированы).

26. CDPC реализует многолетний проект под названием «Искусственный интеллект и уголовно-правовая ответственность в государствах – членах Совета Европы – в случае автоматизированных транспортных средств». Концептуальный документ, лежащий в основе этого проекта, открывается с наиболее актуальных вопросов, а именно: «Кто будет нести ответственность, если полностью автоматизированный автомобиль ранит или убивает человека? При использовании самообучающихся алгоритмов вождения автомобиля, возникает более общий вопрос: как уголовное право должно рассматривать искусственный интеллект? Позже он отмечает, что «правовая база, применяемая в настоящее время к разработке и использованию автоматизированных транспортных средств» базируется на нормативных принципах, разработанных в доцифровую эпоху. Поэтому было бы полезно установить правила, регулирующие любого потенциального преступника. Уголовная ответственность заранее обеспечивает, что в случаях как, автомобильное столкновение или крушение беспилотного летательного аппарата, ни одно государство не столкнётся с неясной правовой ситуацией из-за неподходящих или устаревшей правил. Поскольку потенциальное внедрение автоматизированных транспортных средств повлияет на весь Совет государств – членов Европы и других стран, организация должна сыграть роль в содействии общей разработке принципов, относящихся к развёртыванию ИИ. Необходимо решить несколько вопросов, включая вопрос о том, как различные подходы к тестированию и использованию автоматизированных транспортных средств могут привести к «допустимым рискам», не криминализированным во внутреннем законодательстве (например различные виды использования технологий в автомобилях), а также вопрос о том, что автоматизированному транспортному средству, возможно, в конечном итоге придётся отвечать перед законом как электронное лицо (аналогично корпорациям как юридическим лицам).

### **Гражданская ответственность за ущерб**

27. Если полуавтономное или автономное транспортное средство наносит ущерб, пострадавший может потребовать компенсацию. В случае аварий с участием обычных транспортных средств вред или убытки обычно оцениваются с точки зрения ответственности участников дорожного движения. Действует система ответственности по вине, согласно которой сторона, неосторожно или умышленно

нарушившая правила дорожного движения, обязана возместить ущерб стороне, пострадавшей в результате аварии. Этот тип системы обеспечивает защиту от неисправностей, когда водитель сам не мог предотвратить аварию. В контексте полуавтономных транспортных средств существует больше возможностей для использования этой защиты, когда можно утверждать, что система автоматизации была ответственна за нанесённый ущерб. Кроме того, по отношению к полностью автономным транспортным средствам, система ответственности на основе ошибок будет означать, что пользователю транспортного средства никогда не придётся возмещать ущерб пострадавшей стороне, поскольку эксплуатация транспортного средства полностью вышла из-под их контроля.

28. Для обеспечения того, чтобы потерпевшие стороны могли получить компенсацию за свои убытки многие выступают в пользу системы строгой ответственности. Таким образом, даже при отсутствии доказательств вины владелец или пользователь транспортного средства автоматически несёт ответственность за любой ущерб. В этом случае наличие технологии автоматизированного вождения не повлияет на ответственность владельца. При такой системе от производителя могут потребовать внести часть страховки для каждого транспортного средства, ограничивая при этом свою ответственность за качество продукции. Другой альтернативой может быть система страхования первой стороны, при которой пострадавшие в каждом транспортном средстве получают компенсацию непосредственно от страховщика этого транспортного средства, в то время как немоторизованные участники дорожного движения (пешеходы, велосипедисты и т.д.) по-прежнему защищены ответственностью третьей стороны. В этой системе ущерб, причинённый с помощью вспомогательной технологии вождения будет автоматически компенсироваться страховщиком транспортного средства.

29. Пригодность этих различных режимов будет зависеть от степени автоматизации транспортного средства. Это покажет, насколько справедливо можно ожидать, что водитель и страховщик несут ответственность за ущерб, учитывая степень контроля, которую имел водитель, или должен был иметь над поведением транспортного средства. В Соединённом Королевстве закон об автоматизированных и электрических транспортных средствах 2018 года является примером передовой практики в этой области и включает в себя и исчерпывающий перечень, разъясняющий ответственность владельцев транспортных средств и страховщиков в широком спектре обстоятельств в случае аварии.

### **Ответственность производителя**

30. Поскольку водитель имеет меньший контроль над поведением полуавтономного транспортного средства, чем над обычным автомобилем, часто

возникающий ущерб является ошибкой производителя, повреждение которого происходит по вине производителя, который затем может быть привлечён к ответственности на основании ответственности за качество продукции. Однако технология автоматизированного вождения вызывает ряд особых проблем в связи с действующими правилами ответственности за качество продукции.

31. С точки зрения представления полуавтономных транспортных средств, уже были проблемы с вводящим в заблуждение маркетингом. Например, система Volvo Pilot Assist изначально рекламировалась в разделе «автономное вождение» на сайте компании, несмотря на то, что водители должны были держать руки на руле. Такие ошибки могут ввести в заблуждение клиентов, которые не понимают сложностей этой новой технологии, и производители могли бы быть привлечены к ответственности за любой ущерб, причинённый в результате того, что такие водители полагаются на утверждение о том, что транспортные средства являются автономными. Производители несут ответственность только за продукты, которые считаются дефектными в пределах их разумно ожидаемого использования. По этой причине необходимо учитывать человеческий фактор при тестировании технологий вспомогательного вождения, поскольку некоторые типы неосторожного поведения в определённой степени предсказуемы даже при использовании новых технологий. Производители могут быть привлечены к ответственности за ущерб, причинённый продуктом, если он должен быть адаптирован в соответствии с альтернативными проектами, которые были доступны на момент его продажи. И наоборот, защита «риска развития» означает, что производитель не может быть привлечён к ответственности, если дефект не мог быть известен в то время, когда он был введён в оборот, с учётом уровня технических знаний. Учитывая темпы развития технологий вспомогательного вождения, вполне возможно, что потребителям придётся нести бремя ряда рисков, которые ещё не были научно неизвестны.

32. Кроме того, бремя доказывания в делах об ответственности за продукцию обычно лежит на потерпевшей стороне. Однако сложность технологий вспомогательного вождения делает его невероятно трудным для физического лица или страховщика доказать, что техническая неисправность стала причиной аварии. Это может создать несправедливое бремя для потребителя. Равным образом, степень, в которой производитель может быть привлечён к уголовной или гражданской ответственности в отношении жертв несчастных случаев, может повлиять на их ответственность перед потребителем за продажу дефектного продукта в соответствии с принципом *ne bis in idem*.

## **Страховые проблемы**

33. Страховщики являются ключевыми заинтересованными сторонами в развитии технологий вспомогательного вождения, поскольку автострахование является обязательным. Они имеют право решать, страховать их или нет полуавтономные и автономные транспортные средства, и тем самым определять их коммерческую жизнеспособность. В настоящее время страховщики находятся в сложном положении с точки зрения установления надбавок для водителей полуавтономных транспортных средств, поскольку, хотя эти транспортные средства предназначены для того, чтобы быть более безопасными, чем обычные транспортные средства, на начальных этапах запуска системы неизбежно возникают различные новые риски и опасности. Однако в долгосрочной перспективе средние страховые взносы скорее всего, упадут из-за снижения риска несчастных случаев.

34. Один из способов, с помощью которого страховщики могут решить некоторые проблемы, связанные с установлением страховых взносов и обязательствами, это использование системы оплаты по мере поступления (PAYD) и «чёрных ящиков». PAYD – это модель страхования, использующая телематические системы для расчёта страховых взносов в соответствии с индивидуальным поведением водителя и может быть использована для установления ответственности. Эту технологию можно использовать для наблюдения за салоном автомобиля и обеспечением внимательности водителя даже при использовании вспомогательных средств вождения. Системы PAYD считаются более справедливыми, поскольку с пользователей взимается плата в соответствии с их собственным поведением за рулём, и исследования показали, что они могут положительно повлиять на поведение водителя. Однако, усиление контроля за потребителями таким образом также вызывает вопросы конфиденциальности и безопасности.

## **Конфиденциальность и проблемы кибербезопасности**

35. Полуавтономные и автономные транспортные средства работают с использованием связи между транспортными средствами и транспортными средствами с инфраструктурой, датчиками и картами высокой чёткости. Все это позволяет им учиться у других транспортных средств и обеспечивать максимальную безопасность. Однако, вся эта информация также составляет значительный набор персональных данных, в частности данные о местонахождении транспортного средства (а также его водителя и пассажиров). Внедрение биометрических функций, таких как распознавание отпечатков пальцев, лиц и радужной оболочки глаза, в

системы безопасности транспортных средств и другие системы также подразумевают обработку и хранение личных данных. Есть ещё несколько безответных вопросов относительно информационных систем, используемых в полуавтономных транспортных средствах, например, какой тип информации собирается и почему; кто контролирует и имеет доступ к этой информации; и как долго она хранится? Должны ли данные, записанные автономными транспортными средствами до и во время аварии, автоматически передаваться не только в центральную систему, но и в страховую компанию, полицию и другие регулирующие/правоохранительные органы? В Европейском союзе (ЕС) в этом отношении актуален общий регламент защиты данных (GDPR). Эти правила распространяются на все компании, обрабатывающие данные от субъектов проживающих в ЕС, независимо от места нахождения компании. Необходимо найти правильный баланс между стимулированием инноваций и защитой частной жизни отдельных лиц. Эффективные правила защиты данных, применимые также к контексту подключённых (полу) автономных транспортных средств, будут важной и необходимой частью общего регулирования автономных транспортных средств.

36. Установка компьютера во главе пассажирского транспортного средства вызывает опасения по поводу кибербезопасности. Хакеры потенциально могут взять контроль над автомобилем через беспроводные сети, такие как Bluetooth, системы доступа ключа, сотовые и другие соединения. Опасность взлома будет особенно высока по отношению к транспортному средству 5 уровня, где человек – водитель или, в данном случае, скорее пассажир не обязан контролировать или вмешиваться в любую из задач вождения. Более того, взломанные данные будут финансово ценными и их можно будет продать третьим лицам. Таким образом, обеспечение безопасности информационных систем, используемых технологиях вспомогательного вождения, должно быть приоритетом.

## **Текущее регулирование в Европе**

37. Все государства-члены Совета Европы имеют свои режимы регулирования автономных и полуавтономных транспортных средств. Многие страны, такие как Германия и Великобритания, разработали специальные законы о технологиях вспомогательного вождения, в то время как другие применяют существующие законы для обычных транспортных средств к полуавтономным транспортным средствам. Тем не менее, существует также ряд важных международных и региональных документов. В области определения требований безопасности транспортных средств можно утверждать, что относительная важность конкретных национальных законов быстро снижается, особенно в рамках государств Европейского Союза.

38. Как отметил г-н Шамп, в настоящее время уже существует «множество законов» на различных уровнях юрисдикции, касающиеся дорожных транспортных средств. Этот закон охватывает целый ряд вопросов, включая страхование, ответственность за качество продукции, международные стандарты транспортных средств, уголовную ответственность за правонарушения, связанные с вождением, гражданско-правовые санкции за нарушение правил вождения, потребительскую информацию и маркетинговые стандарты, водительские права, процедуры расследования несчастных случаев, защита данных, регулирование рынков такси и частного проката, регулирование транспортных средств общего пользования, общие правила дорожного движения и т.д. Этот закон был разработан в отношении неавтономных транспортных средств в виду, или, в лучшем случае тех, у кого есть некоторые функции помощи водителю, под постоянным контролем ответственного человека. Это разделение между водителем и транспортным средством в настоящее время исчезает, и новые законы и нормативные акты должны будут решить, должны ли гарантии безопасности для автоматизированного вождения сосредотачиваться на водителе или транспортном средстве. В некоторых случаях сама технология в значительной степени решит эту дилемму, особенно когда речь заходит о полностью автономных транспортных средствах без участия пользователя (или, в лучшем случае с удалённым диспетчером, способным принимать решения в исключительных обстоятельствах).

39. На международном уровне Венская конвенция о дорожном движении 1968 года посвящена общему праву дорожного движения. Она была подписана и ратифицирована 38 государствами – членами Совета Европы. Статьи 8.5 и 13 гласят, что все водители должны уметь управлять своим автомобилем в любое время. В 2016 году в Статью 8 конвенции был добавлен новый абзац, предусматривающий создание автоматизированных транспортных средств. В результате автономные и полуавтономные транспортные средства будут считаться соответствующими конвенции при условии, что система может быть заблокирована водителем или соответствовать требованиям Соглашения европейской экономической комиссии Организации Объединённых Наций (ЕЭК ООН) 1958 года и Соглашению о глобальных технических правилах (ГТП) 1998 года.

40. Соглашение ЕЭК ООН 1958 года было заключено с целью облегчить принятие единых условий официального утверждения автомобильного оборудования и запчастей по всей Европе. Соглашение о ГТП 1998 года было разработано для создания глобального процесса утверждения этого типа, который был бы применим во всех странах мира. На уровне ЕС была принята директива, гарантирующая, что после сертификации транспортных средств и их компонентов в одном государстве-члене, они не могут быть исключены с рынков других государств, если нет достаточных доказательств того, что это может серьёзно угрожать безопасности дорожного движения.

41. В 2018 году ЕЭК ООН учредила специальную рабочую группу по автономным и подключённым транспортным средствам (GRVA). Эта группа вместе с гражданским обществом помогает мобилизовать опыт из ключевых отраслей промышленности для реализации видения новой устойчивой мобильности и поддержки массового внедрения автономных транспортных средств на дорогах. Существует также целый ряд неофициальных рабочих групп (НРГ) по конкретным вопросам безопасности транспортных средств. НРГ по автоматизированной командной функции рулевого управления (ACSF) предложила поправки к Правилу 79 ЕЭК ООН с целью включения положений о технологии ACSF. Принятые поправки включают то, что в случае вмешательства ACSF водитель должен быть в состоянии получить контроль над транспортным средством в течение четырёх секунд; а также требуется немедленное отключение и транспортное средство должно быть способно оставаться в полосе движения, сохранять дистанцию и справляться со сценариями удара сзади на автомагистралях.

42. ЕС также принял документы регулирующие присвоение риска в связи с использованием автотранспортных средств. Двумя основными законодательными актами, регулирующими ответственность, являются Директива по автомобильному страхованию и Директива об ответственности за продукцию. Основные правила об ответственности за ущерб, причинённый в результате дорожно-транспортных происшествий не согласованы на уровне ЕС. Директива по автострахованию предписывает только минимальное страхование ответственности перед третьими лицами. Автономные транспортные средства подпадают под определение «транспортного средства», содержащееся в статье 1 настоящей Директивы, и поэтому автоматически попадают под её положения. Директива об ответственности за продукцию устанавливает правила, касающиеся ответственности производителя и права потребителей. Они основаны на режиме безошибочной ответственности, означающий, что производитель бракованного товара должен предоставить компенсацию за телесные повреждения, причинённые его продуктом независимо от халатности физического лица. Презентация продукта, разумно ожидаемое использование продукта и время, когда он был выложен на рынок, всё влияет на определение того, является ли он «дефектным». Существует ограниченный перечень отступлений, которые могут исключить ответственность за продукцию. Однако данная Директива распространяется только на ответственность производителей некачественной продукции, которая может оказаться недостаточной, чтобы иметь дело с ответственностью производителя за травмы, вызванные автономным и полуавтономными транспортными средствами. Кроме того, при таком режиме работы стоимость научно неизвестных рисков будет взвалена на плечи потерпевшей стороны.

43. Г-жа Хамсен рассказала нам о подходе к регулированию автономных транспортных средств и систем, применяемом властями Германии, европейской страны с крупнейшим сектором производства транспортных средств. С 2017 года

законодательство о безопасности дорожного движения позволило достичь 3 уровня автоматизации, который в некоторых случаях не требует вмешательства водителя. Этот закон был полезен для автомобильной промышленности, поскольку обеспечивал ясность и безопасность для инвестиций. Законодательство чётко определяет, когда водитель – человек может быть подвергнут «адаптированным» обязательствам по обращению внимания к трафику, и на каких предварительных условиях. Если условия не будут выполнены, водитель остаётся ответственным в случае аварии. Технические предпосылки включают в себя систему, способную уважать нормальные правила дорожного движения и правила, которые водитель-человек обязан соблюдать. Система должна быть описана в международных правилах, действующих в Германии, или в заявке на освобождение в соответствии с законодательством ЕС. Пока ещё нет ни одного применимого международного правила, касающиеся автономных транспортных средств; ожидается, что GRVA (см. выше) внесёт предложения, но, по словам г-жи Хамсен, скорее всего, не завершит свою работу до следующего года. Если в техническом описании системы производителя указано, что транспортное средство имеет 3 уровень автоматизации и может соответствовать правилам дорожного движения, то его водителю разрешается передать управление автомобилем автономной системе транспортного средства, хотя водитель всё ещё должен иметь возможность возобновить управление, если предварительные условия для использования системы автономного транспортного средства больше не будут выполняться.

## **Выводы и рекомендации**

44. Использование полуавтономных транспортных средств, вероятно, значительно увеличится в ближайшие годы, поскольку технология становится все более совершенной и надёжной. Некоторые даже считают возможным, что в течение следующего десятилетия будет разработан полностью автономный автомобиль. Такие транспортные средства имеют потенциал для трансформации личной мобильности и резкого улучшения безопасности дорожного движения. На этом этапе будет технически возможно создать автономные транспортные средства, которые никогда не будут «виноваты» в аварии (за исключением случаев неисправности) и которые даже смогут избежать большинства или всех столкновений, которые были бы «виной» другого транспортного средства (или его водителя), либо пешехода или велосипедиста. В мире, где не все автомобили будут автономными, автономные транспортные средства, запрограммированные на предотвращение любых столкновений, не смогут продвигаться очень быстро, поскольку они будут регулярно «отсекаться» от других, неавтономных транспортных средств, под управлением человека. И последнее, но не менее важное: когда автономные автомобили будут намного безопаснее даже чем самый лучший и

внимательный водитель, должны ли законодатели запрещать неавтономные транспортные средства? Будет ли у них даже позитивное обязательство сделать это, чтобы защита права на жизнь в соответствии со статьёй 2 Европейской конвенции о правах человека (ETS № 5)?

45. Поскольку транспортные средства регулярно пересекают границы, будущее регулирование автономных транспортных средства должно разрабатываться с учётом глобального видения. При этом должны учитываться перспективы всех регионов и работу, проводимую в различных контекстах по всему миру. Различные международные органы, каждый из которых работает в своей области и специализации, должны учитывать деятельность друг друга для того, чтобы обеспечить гармоничный и всесторонний общий результат. Тем, кто занимается регулированием автономных транспортных средств, следует обратить особое внимание на общие принципы регулирования и правовые рамки, разрабатываемые для ИИ, и провести специальную оценку воздействия автономного транспортного средства на права человека, чтобы предвидеть проблемы, которые могут возникнуть.

46. Совет Европы уже внимательно относится к аспектам автономного вождения, которое подпадает под его институциональный мандат. Концентрируясь на уголовно-правовых аспектах автономных транспортных средств, в частности, на влиянии искусственного интеллекта на права человека в целом, организация играет на его сильных сторонах и может внести важный вклад в продолжающиеся дебаты. Одна из целей моего доклада – внести вклад в этот процесс. Поэтому я предлагаю Ассамблее подчеркнуть важность разработки и внедрения автономных транспортных средств как одну из областей, в которых системы искусственного интеллекта могут оказать особое влияние на права человека и демократические общества. Это следует принять во внимание САНАІ при составлении и учёте карт рисков и возможностей искусственного интеллекта, а также необходимости обязательной правовой базы для регулирования её разработки и эксплуатации.

47. Дополнительные рекомендации содержатся в прилагаемом проекте резолюции и рекомендации.

## Китайские взгляды на аналитику больших данных



DEREK GROSSMAN, CHRISTIAN CURRIDEN, LOGAN MA,  
LINDSEY POLLEY, J.D. WILLIAMS, CORTEZ A. COOPER III

# Chinese Views of Big Data Analytics



ДЕРЕК ГРОССМАН, КРИСТИАН КЕРРИДЕН, ЛОГАН МА, ЛИНДСИ ПОЛЛИ,  
Д. УИЛЬЯМС, КОРТЕЗ А. КУПЕР III

## **Ограниченные права на печатное и электронное распространение.**

Этот документ и содержащиеся в нём товарные знаки защищены законом. Отмеченная в нём интеллектуальная собственность RAND предоставляется только для некоммерческого использования.

Неавторизованное размещение данной публикации в Интернете запрещено. Разрешено дублировать этот документ только для личного использования, если он не изменён и не полон. Разрешение требуется от RAND для воспроизведения или повторного использования в другой форме любых исследовательских документов для коммерческого использования. Для получения информации о разрешениях на перепечатку и размещение ссылок посетите [www.rand.org/pubs/permissions](http://www.rand.org/pubs/permissions).

RAND Corporation – исследовательская организация, которая разрабатывает решения для общественных политических вызовов, призванных помочь сделать сообщества во всем мире более зажиточными, безопасными и защищёнными. RAND является некоммерческой, беспристрастной и приверженной общественным интересам организацией. Публикации RAND не обязательно отражают мнение клиентов и спонсоров исследований.

### Поддержка RAND

Сделайте не облагаемый налогом благотворительный взнос на [www.rand.org/giving/contribute](http://www.rand.org/giving/contribute), [www.rand.org](http://www.rand.org)

## **Предисловие**

Стремление Китая наделить искусственный интеллект способностью выполнять разнообразие гражданских и военных функций начинается с освоения аналитики больших данных с использованием компьютеров для анализа больших наборов данных.

Настоящий доклад содержит предварительный анализ ожиданий Китая в отношении планов и стратегий развития аналитики больших данных, задействованных правительственными учреждениями, а также некоторые конкретные приложения доклада для работы с большими данными, которые он использует. Этот доклад предназначен для представления начального базового описания усилий Китая в этой области и определения области для дальнейшего

углублённого исследования. В докладе используются китайские первоисточники, чтобы понять текущее состояние национальной стратегии Пекина по работе с большими данными и последствия её реализации.

Это исследование спонсировалось правительством США и проводилось в Центре кибернетической политики и разведки Исследовательского отдела национальной безопасности РЭНД, который управляет Национальным научно-исследовательским институтом обороны (NDRI), финансируемым из федерального бюджета Центра исследований и разработок при поддержке Управления Министра обороны, Объединённого штаба, Объединённого боевого Командования, ВМФ, Корпуса морской пехоты, Военной разведки.

Для получения дополнительной информации о RAND Cyber and Intelligence Policy Centre, см. [www.rand.org/nsrd/intel](http://www.rand.org/nsrd/intel) или свяжитесь с директором (контактная информация указана на сайте).

## Резюме

Стремление Китая к созданию возможностей искусственного интеллекта (ИИ) для выполнения разнообразных гражданских и военных функций начинается с освоения *аналитики больших данных* – использования компьютеров для понимания больших наборов данных.

Наши исследования показывают, что Китай агрессивно стремится стать мировым лидером в области анализа больших данных. В ходе 19-го Национального съезда Коммунистической партии Китая в октябре 2017 года председатель КНР Си Цзиньпин подчеркнул необходимость «содействия более углублённой интеграции Интернета, больших данных и искусственного интеллекта в реальную экономику»<sup>1</sup>.

Усилия Пекина руководствуются национальной стратегией больших данных, которая включает в себя экономические, военные, полицейские и разведывательные функции. В «Плане действий по содействию развитию больших данных» Государственного совета Китайской Народной Республики от 2015 года частично описан общий подход Китая к развитию и использованию аналитики больших данных. Значительные усилия по работе с большими данными прилагают Министерство промышленности и информационных технологий, Комиссия по национальному развитию и реформам, Министерство образования, Министерство общественной безопасности, Народно-освободительная армия (НОАК) и Министерство науки и технологий.

---

<sup>1</sup> “Full Text of Xi Jinping’s Report at the 19th CPC National Congress,” Xinhua, November 3, 2017.

Наше исследование, основанное в основном на китайских первоисточниках, указывает, что Пекин использует большие данные, в целях достижения Китаем статуса великой державы.

Помимо анализа этих первоисточников, мы оценили конкретные планы и рекомендации на уровне министерств. Так, Министерством промышленности и информационных технологий с помощью RAND, обладающем глубокими познаниями в области прогнозирования развития будущего, исследовались сведения об использовании аналитики больших данных в СМИ – государственных официальных и полуофициальных китайских изданиях, редакционных статьях, научных журналах и соответствующих западных публикациях, посвящённых китайским теоретическим и практическим интересам в области аналитики больших данных в целях поиска возможных направлений движения Пекина в развитии и применении искусственного интеллекта. Этот анализ привёл к выявлению и сосредоточению внимания на двух чётких приоритетных областях для гражданских и военных лидеров: использование Пекином аналитики больших данных для внутреннего наблюдения за своим населением, и его использование для повышения своего военного потенциала.

На внутренней арене цель Коммунистической партии Китая состоит в использовании возможности анализа больших данных, чтобы строго и всесторонне контролировать население Китая. Силы общественной безопасности Китая, пожалуй, с наибольшим энтузиазмом взяли на вооружение аналитику больших данных. В отсутствие законов, ограничивающих сбор данных в целях общественной безопасности, такие мощные инструменты значительно улучшают работу Министерства общественной безопасности. Служба безопасности имеет возможность быстро связывать судимости с практически любыми другими данными, которые считаются имеющими отношение к задержанию предполагаемых преступников. Более того, улучшения в аналитике больших данных напрямую поддерживают строгий мониторинг и контроль над гражданами Китая – включая и этнические меньшинства с помощью любых признаков предполагаемой нелояльности или других проступков. Национальная социальная Кредитная система – является репутационной онлайн-системой, которая позволяет или ограничивает гражданам доступ к социальным благам на основании оценённого поведения, планируется, что она появится в сети к 2020 году. Китай пытается создать систему больших данных национального уровня, способную буквально предотвращать преступления и политическое инакомыслие до того, как это произойдёт. Пекин будет делать это, отслеживая поведение, привычки, ассоциации и чувства каждого члена общества, чтобы оценить склонность каждого человека к участию в такой деятельности.

В военной сфере Пекин проявляет особый интерес к использованию больших данных и в конечном итоге, ИИ для улучшения самых разнообразных возможностей

PLA. Китайская НОАК надеется освоить использование больших данных национальной обороны в достаточной степени, чтобы контролировать, или даже доминировать в информационной среде ведения войны, в которой контроль информации является ключевым фактором против великих держав. Некоторые из основных областей, представляющих интерес для применения больших данных в вооружённых силах, включают командование, управление, связь, компьютеры, разведку, наблюдение и оборудование для разведки и техническое обслуживание; логистику; здравоохранение; мобилизацию; подготовку; набор персонала; моделирование и симуляцию ситуаций, а также кибер-безопасность. Использование больших данных в сфере национальной обороны также способствует долгосрочным усилиям Пекина, о которых президент Си заявил в своей речи на 19-м съезде партии, чтобы к 2030 г.<sup>2</sup> стать глобальным центром ИИ.

Дальнейшее исследование по следующим вопросам позволят больше понять китайское участие и инвестиции в аналитику больших данных:

Как Китай может использовать развитие своих возможностей в применении больших данных для мониторинга и контроля своего населения?

Какую позицию занимает развивающийся потенциал Китая в области анализа больших данных по сравнению с США в военной сфере?

Как Китай использует анализ больших данных в поддержку своих исследований и процессов их разработки?

Как Китай оценивает технологические или экономические тенденции, используя аналитику больших данных?

Как китайское руководство планирует реализацию государственных услуг на основе больших данных и их последствиях для внутренней стабильности?

## **Благодарности**

Мы хотим поблагодарить наших спонсоров в правительстве США в их стремлении глубже понять китайские подходы к искусственному интеллекту и аналитике больших данных. Мы также хотели бы поблагодарить наших рецензентов и коллег из корпорации RAND Тимоти Р. Хиту и Ларри Харрисона за их вдумчивые комментарии. В заключение, мы хотели бы поблагодарить Ричарда С. Гирвена и Сину Бигли из RAND Cyber and Intelligence Policy Center за их поддержку и руководство.

---

<sup>2</sup> “Full Text of Xi Jinping’s Report at the 19th CPC National Congress,” 2017.

**Сокращения**

AI	искусственный интеллект
C4ISR	команда, контроль, связь, компьютер, разведка, наблюдение и разведка
CCP	Коммунистическая партия Китая
EM	электромагнитный
IJOP	Интегрированная платформа совместных операций
МИПТ	Министерство промышленности и информационных технологий
MPS	Министерство общественной безопасности
NDRC	Национальная комиссия по развитию и реформам
OPM	Управление персонала США
PGIS	Географическая информационная система полиции
PLA	Народно-освободительная армия
PLAAF	ВВС Народно-освободительной армии
PLAN	Флот Народно-освободительной армии
PLASSF	Силы стратегической поддержки Народно-освободительной армии
R&D	исследования и разработки

## Глава первая

### Вступление

Высшее руководство Китая часто выражает мнение, что продолжающийся экономический и военный рост неразрывно связан с приобретением и собственной разработкой передовых технологий. Можно отметить, что практически во всех областях развития науки и технологий, будь то робототехника, генетика, космические технологии, дроны, нанотехнологии, фармацевтика, микропроцессоры или солнечная энергия, Китай неизменно занимает лидирующие позиции в мире или приближается к ним. Однако Пекин признает, что ценность только обладания передовыми технологиями ограничена. Таким образом, Китай ищет способы соединить и снабдить свои передовые технологии данными, в рамках системного подхода, чтобы раскрыть весь потенциал новых технологий<sup>3</sup>.

Следующим и, возможно, окончательным шагом будет развитие и внедрение автоматизированного процесса принятия решений, широко известного как искусственный интеллект (ИИ) в широкий спектр правительственных, экономических и военных приложений<sup>4</sup>. Действительно, в последние два года интерес Китая в области технологий искусственного интеллекта занял центральное место<sup>5</sup>. В июле 2017 года Государственный совет КНР объявил, что Китай «стремится к 2020 году идти в ногу с ведущими странами в области технологий искусственного интеллекта. Более того, Китай планирует стать «крупным центром инноваций в области ИИ и ведущим мировым лидером в области технологий и приложений ИИ к 2030 году»<sup>6</sup>.

Чтобы стать глобальным центром искусственного интеллекта Китаю необходимо сначала усовершенствовать промежуточный и ключевой шаг вперед, называемый аналитикой больших данных.

---

<sup>3</sup> Подробнее по этой теме см. Джеффри Энгстрем, Системное противостояние и война за уничтожение систем: Как китайская Народно-освободительная армия стремится вести современную войну, Санта-Моника, Калифорния: Корпорация RAND, RR-1708-OSD, 2018.

<sup>4</sup> Конечно, универсального определения ИИ нет, но наиболее общепринятым является компьютерное принятие решений. Для получения дополнительной информации см. Келли М. Сайлер, Искусственный интеллект и национальная безопасность, Вашингтон, округ Колумбия: Исследовательская служба Конгресса, R45178, обновлено в ноябре 21, 2019.

<sup>5</sup> Для получения дополнительной информации о недавнем фокусе Китая на ИИ, см. Грэм Уэбстер, Рогиер Кримерс, Пол Триоло и Эльза Каня, "План Китая" Вести "в ИИ: цель, перспективы и проблемы", Новая Америка, запись в блоге, 1 августа 2017 года.

<sup>6</sup> "China Maps Out AI Development Plan", Синьхуа, 20 июля 2017 года. Также смотрите "План развития искусственного интеллекта нового поколения", Госсовет, 20 июля 2017 года.

Согласно IBM, аналитика больших данных – это процесс, с помощью которого компьютеры оценивают большие наборы данных, чтобы позволить людям (а не компьютерам, каковым является ИИ) «получать новые идеи, приводящие к более качественным и быстрым решениям». Аналитика больших данных включает в себя такие возможности, как «текстовая аналитика, машинное обучение, прогнозная аналитика, интеллектуальный анализ данных, статистика и обработка естественного языка»<sup>7</sup>.

Наше исследование показало, что Китай активно работает над тем, чтобы стать мировым лидером в области анализа больших данных. Во время 19-го съезда Компартии в октябре 2017 года председатель Китая Си Цзиньпин заявил, что Китаю необходимо «способствовать более глубокой интеграции Интернета, больших данных и искусственного интеллекта с реальной экономикой»<sup>8</sup>. Китай чаще всего переводил большие данные как 大数据 (da shuju) – термин, который не только соответствует дословному переводу английского термина, но и сохраняет то же значение в китайской системе. Обсуждения в материалах китайских первоисточников подтверждают, что концепция больших данных в Китае практически такая же, как и в Соединённых Штатах Америки. И, как и на Западе, термины «большие данные» и «ИИ» (人工智能, rengong zhineng) иногда используются как синонимы, предполагая, что путаница по поводу их точных определений также может существовать в китайской системе, хотя заявления с самых высоких уровней, таких как Си Цзиньпин, как уже упоминалось ранее, указывает на то, что китайское руководство действительно понимает разницу.

Цель этого отчёта – выявить и оценить ожидания Китая в отношении аналитики больших данных. Чего конкретно хотят добиться китайцы? Что, по их мнению, возможно? Кроме того, в каких сферах общества они планируют применять аналитику больших данных? Наше исследование указывает на множество видов деятельности, которые, как надеется Пекин, выиграют от внедрения приложений для работы с большими данными. Для этого отчёта мы рассмотрели два тематических исследования, выбранных нашим спонсором.

Первый – это взгляд на применение Китаем аналитики больших данных для предотвращения преступности или, что более цинично, для мониторинга и контроля своего населения.

Во втором изучаются планы Пекина по повышению военного развития Народно-освободительной армии (НОАК) и планировании операций с помощью

---

<sup>7</sup> IBM, "Аналитика больших данных", веб-страница, без даты. Методы анализа больших данных можно разделить на семь ключевых областей: изучение правил ассоциации, анализ дерева классификации, генетические алгоритмы, машинное обучение, регрессионный анализ, анализ настроений и анализ социальных сетей. Для получения дополнительной информации см. Preet Navdeep, Manish Arora и Neeraj Sharma, "Роль аналитики больших данных в анализе проектов электронного управления", Gian Jyoti e-Journal, том 6, № 2, апрель-июнь 2016.

<sup>8</sup> "Полный текст доклада Си Цзиньпина на 19-м Национальном конгрессе КПК", Синьхуа, 2017.

аналитики больших данных. Мы также определили дополнительные тематические исследования по применению больших данных для практики исследований и разработок (НИОКР), операций разведки и контрразведки или оказания государственных услуг населению.

## Источники и методы

Мы сосредоточили наши исследования и анализ на китайских первичных материалах. К ним относятся стратегические планы и руководящие документы высокого уровня, такие как «План действий Государственного совета 2015 года по содействию развитию больших данных [促进大数据发展行动纲要]» и «Список 2018 года Национальной комиссии по развитию и реформам» (NDRC). Интернет +, AI и цифровая экономика экспериментируют с получателями основных проектов»<sup>9</sup>. Также были оценены конкретные планы на уровне министерств и рекомендации Министерства промышленности и информационных технологий (МИИТ). Кроме того, мы изучили сводки государственных СМИ, таких как People's Liberation Army Daily (解放军报), о том, как Китай использует аналитику больших данных. Мы также исследовали менее известные официальные и полуофициальные источники информации, такие как Сети по закупкам вооружений и техники НОАК [全军武器装备采购信息网]. Также изучались неофициальные китайские публикации и обсуждения, особенно те, которые появляются в академических журналах и в качестве передовых статей в литературе. Однако наше внимание к китайским первичным источникам не исключало рассмотрения соответствующих западных публикаций, которые освещают теоретические и практические интересы Китая в области анализа больших данных. Наконец, мы привлекли исследователей корпорации RAND, обладающих глубокими познаниями о будущем вычислительной техники, чтобы помочь нам наметить потенциальные направления, по которым Пекин может следовать вперёд, продвигая аналитику больших данных и искусственный интеллект.

---

<sup>9</sup> NDRC, "List of 2018 Internet +, AI, and Digital Economy Experiments Major Projects Recipients [2018年“互联网+”、人工智能创新发展和数字经济试点重大工程拟支持项目名单公示]," December 27, 2017.

## **Организация этого отчёта**

Во второй главе более подробно будет рассмотрена деятельность руководства Китая по аналитике больших данных, а также агентства, отрасли и корпорации, которые участвуют в этой области. Также мы определим проекты или инициативы, которые находятся в стадии реализации, и основные ожидаемые результаты этих усилий. В третьей главе мы рассмотрим первый пример того, как Китай использует или стремится использовать аналитику больших данных для внутренней безопасности, а также для мониторинга и контроля своего населения. В четвертой главе рассмотрим наш второй практический пример потенциальной полезности аналитики больших данных для PLA. В пятой главе предлагаются заключительные выводы и выделяются области, которые могут быть полезны для дополнительных исследований и анализа.

## **Глава вторая**

### **Китай уделяет приоритетное внимание аналитике больших данных**

Во время своей марафонской речи на 19-м съезде Компартии в октябре 2017 года председатель Си заявил, что Китаю необходимо «реализовать более глубокую интеграцию Интернета, больших данных и искусственного интеллекта с реальной экономикой»<sup>10</sup>. В этом заявлении, как и в других своих выступлениях, Си Цзиньпин чётко обозначил то внимание, которое Китай уделяет возможности в достижении анализа больших данных. Кроме того, упомянув в одном предложении большие данные и ИИ, Си подчеркнул тесную связь между этими концепциями в мировоззрении Китая и тем, как сочетание этих двух факторов приведёт Китай к созданию общества, основанного на ИИ, к 2030 году. С исторической точки зрения заявление председателя Си является кульминацией в обсуждении и озвучивании стратегических планов высшего руководства страны за последние нескольких лет о необходимости развития аналитического потенциала Китая в области больших данных. В этой главе более подробно рассматриваются эти документы и выделяются несколько основных инициатив в области больших данных, осуществляемых в настоящее время, а также их ожидаемые результаты. Мы также изучаем деятельность различных китайских государственных учреждений, занимающихся аналитикой больших данных.

---

<sup>10</sup> "Full Text of Xi Jinping's Report at the 19th CPC National Congress," 2017.

## **Реализация национальной стратегии больших данных**

В декабре 2017 года Си Цзиньпин председательствовал на втором коллективном учебном заседании Центрального политбюро Коммунистической партии Китая (ССР), на котором обсуждалась реализация «национальной стратегии больших данных» Пекина. На заседании Си Цзиньпин сказал, что Китай должен «стремиться к взятию на себя инициативы по углублённому пониманию текущего состояния и тенденций развития больших данных и их влияния на экономическое и социальное развитие»<sup>11</sup>.

Позже в своём выступлении Си Цзиньпин разъяснил, что социальное развитие означает, что Китай должен «улучшать жизнь людей» с помощью больших данных. Си Цзиньпин также заявил, что «мы должны максимально использовать платформу больших данных для всестороннего анализа факторов риска, чтобы улучшить способность распознавать, прогнозировать и предотвращать эти факторы риска». Хотя мы не смогли найти ни одного документа, который бы однозначно соответствовал национальной стратегии по развитию исследований больших данных, а комментарии Си Цзиньпина чётко указывают на то, что Китай придерживается единого подхода к достижению своих целей в этой области. Ключевая тема, по-видимому, заключается в том, что по мере совершенствования технологий Пекин будет все чаще использовать компьютерный, а не человеческий анализ тенденций, чтобы увеличить эффективность государственной деятельности. Ключевые компоненты китайской стратегии представлены в планах, обсуждаемых в этой главе.

## **Заинтересованные стороны в национальной стратегии больших данных**

Прежде чем искусственный интеллект занял центральное место в 2017 году, была увеличена активность в области стратегии больших данных, особенно с 2014 по 2015 год. В число вовлечённых структур входили Государственный совет и МИИТ. Кроме того, в 2018 году NDRC объявил о государственной поддержке конкретных проектов, связанных с большими данными, а Министерство образования поручило подчинённым учреждениям подготовиться к обучению будущих кадров. Хотя мы знаем, что Министерство общественной безопасности Китая (MPS) и PLA используют аналитику больших данных для выполнения своих задач, мы не смогли найти никаких руководящих документов по большим данным

---

<sup>11</sup> «Китай: Си Цзиньпин председательствует на второй коллективной сессии Политбюро по национальной стратегии больших данных», Синьхуа, 9 декабря 2017 года. Резюме можно найти в "Китай должен ускорить реализацию стратегии больших данных: Си", Синьхуа, CHR2017120927141422, 9 декабря 2017 года.

или аналитике больших данных от этих учреждений. Конечно, в реализации национальной стратегии Китая по работе с большими данными участвуют и другие государственные организации и комитеты, включая, например, Министерство науки и технологий, но мы также не смогли найти существенной информации о других учреждениях, участвующих в реализации стратегии.

## Государственный совет

В 2015 году Государственный совет Китая опубликовал «План действий по содействию развитию больших данных»<sup>12</sup>. Согласно плану, Китай имеет преимущество в области больших данных и должен использовать его для создания новых платформ для обмена данными с общественностью и расширения возможностей правительства предоставлять услуги населению. Согласно Плану действий Государственного совета, например, отмечается призыв к созданию баз данных о населении Китая, корпорациях и рабочих единицах, природных ресурсах и географии к 2018 году. Хотя мы не смогли оценить прогресс во всех областях, но обнаружили, что были достигнуты значительные успехи в интеграции и представлении географических данных<sup>13</sup>.

В рамках плана действий Пекин также стремится создать совместные системы медицинских информационных данных, платформу управления образованием, базу данных по транспорту и туризму, которая объединяет данные о транспорте, общественной безопасности, погоде, землетрясениях и географические данные.

Кроме того, Государственный совет поручил всем органам власти на уровне городов и выше внедрить приложение по связям с правительством и государственным услугам с целью использования этих новых систем для укрепления законности, сокращения коррупции и улучшения управления рисками. Совет специально призвал к созданию механизмов для отслеживания экономических данных, включая финансирование, налоги, аудит, потребление, инвестиции, импорт и экспорт, строительство, занятость, распределение доходов, состояние электроэнергетики и промышленности, качество и безопасность продукции, а также энергосбережение. Госсовет заявил, что предусматривает создание умных городов, которые будут постоянно собирать данные о гражданах Китая для улучшения услуг и, что ещё более гнусно, для мониторинга и контроля за ними.

---

<sup>12</sup> State Council, "Action Plan for Promoting the Development of Big Data [促进大数据发展行动纲要]," August 31, 2015.

<sup>13</sup> See, for example, Central People's Government [中华人民共和国中央人民政府], "Our Nation Will Build a "Unified Picture" Government Affairs Geographic Information Big Data System [我国将建设政务地理信息大数据 "一张图"]," December 26, 2017.

Госсовет также поручил подведомственным организациям изучить основные тенденции данных сельскохозяйственного сектора. Совет хочет использовать аналитику больших данных для мониторинга производства, потребления, хранения, импорта и экспорта, цен, сведений о затратах на сельскохозяйственную продукцию и других показателей состояния сельскохозяйственного сектора. Это примечательно из-за высокого уровня чувствительности сельского хозяйства в Китае. Китайские лидеры, исторически сталкивающиеся с угрозами своему продолжающемуся правлению из-за социальных волнений в сельских районах, делают упор на возможность предотвратить и интегрировать этот тип потрясений в современную эпоху.

С планом использования больших данных связан отдельный план Государственного совета ССР по ИИ на 2017 год: «План развития искусственного интеллекта нового поколения [新一代人工智能发展规划]». В плане отмечалось, что ИИ стал новым центром международной конкуренции и что достижение преимущества в этой технологии имеет решающее значение для поддержания национальной конкурентоспособности и безопасности. Госсовет заявил, что Китай должен создать базовую инфраструктуру ИИ, в том числе усовершенствовать инфраструктуру больших данных, это шаг, демонстрирующий, что Пекин рассматривает большие данные как важный строительный блок для достижения своих целей в области ИИ к 2030 году.

## **Министерство промышленности и информационных технологий**

В 2016 году МИТ опубликовал свой собственный план по большим данным «План развития индустрии больших данных Министерства промышленности и информационных технологий (2016–2020 гг.)»<sup>14</sup>. По прогнозам МИТ, в период с 2016 по 2020 год доходы от продуктов и услуг, связанных с большими данными, достигнут 160 миллиардов долларов для китайской индустрии больших данных, включая как частный сектор, так и государственные предприятия, и будут расти примерно на 30% в год. МИТ сосредоточился в первую очередь на централизации китайских наборов данных, заявив, что создаст национальный центр обработки данных для управления обществом, улучшения государственного управления и повышения уровня жизни людей.

МИТ также отметил, что правительство будет помогать компаниям, работающим с большими данными, получать финансирование через государственно-частные партнёрства, поощряя слияния и поглощения.

---

<sup>14</sup> МИТ, “Ministry of Industry and Information Technology Big Data Industry Development Plan (2016–2020) [工业和信息化部关于印发大数据产业发展规划 (2016–2020年)],” December 18, 2016.

МИТ поощряет интеллектуальное производство, а также обмен данными между телекоммуникационными и интернет-компаниями, финансовыми, промышленными, общественными, транспортными и медицинскими компаниями. Решающее значение имел призыв МИТ к стандартизации качества данных для облегчения эффективного обмена данными между китайскими организациями, занимающимися большими данными.

### Национальная комиссия развития и реформ

Согласно её вебсайту, основная задача NDRC – «формулировать и воплощать стратегии национального экономического и социального развития, годовые планы, среднесрочные и долгосрочные планы развития» для дальнейшего представления Всекитайскому собранию народных представителей «от имени Государственного совета»<sup>15</sup>. В 2017 году NDRC опубликовала свой полный список награждённых проектов больших данных в «Списке основных проектов 2018 в области Интернета+, искусственного интеллекта и цифровой экономики»<sup>16</sup>. В этот список вошли шесть проектов в области судоходства и логистики, пять по распознаванию лиц, пять по облачным вычислениям, четыре по торговле, два по обработке естественного языка, два по обмену данными о здоровье и один по созданию сервисной платформы для источников производства энергии в Китае. Мы ожидаем, что награждённые NDRC сыграют важную роль в развивающемся мире больших данных и искусственного интеллекта в Китае.

### Министерство образования

План действий Госсовета в отношении больших данных побудил другие китайские правительственные учреждения (например, Министерство образования) издать свои собственные инструкции для подчинённых по выполнению плана. В апреле 2018 года Министерство образования опубликовало: «План действий по инновациям в области искусственного интеллекта для колледжей и университетов»<sup>17</sup>. В руководстве образовательным учреждениям было предложено «продвигать исследования фундаментальных теорий в такой области, как анализ больших данных». Кроме того, Министерство стремилось «поощрять корректировки и интеграцию между такими академическими областями, как наука и технологии искусственного интеллекта наука о данных, и технология больших данных». В руководстве также обозначено несколько ключевых областей исследования в

---

<sup>15</sup> NDRC, "Main Functions of the NDRC," undated.

<sup>16</sup> National Reform and Development Commission, 2017.

<sup>17</sup> Ministry of Education, "Ministry of Education Issues and Distributes the Artificial Intelligence Innovation Action Plan for Colleges and Universities [教育部印发《高等学校人工智能创新行动计划》," April 15, 2018.

области анализа больших данных, включая здравоохранение, городское планирование и финансовые потоки, с акцентом на визуальный анализ на основе больших данных. По заявлению Министерства, анализ больших данных уже является частью учебной программы в китайских профессиональных училищах, что подчёркивает большое значение, которое Пекин придаёт надлежащей подготовке своих кадров для проведения этой деятельности.

## **Истинный общеправительственный подход**

Наше исследование показывает, что национальная стратегия Китая по работе с большими данными включает в себя несколько государственных организаций, что делает эту стратегию, аналогичной стратегии в отношении ИИ и поистине общегосударственной задачей. Руководителями стратегии, по-видимому, являются Государственный совет, NDRC и МИТ, возможно, в том же порядке. Другие организации, такие как Министерство образования, поддерживают национальную стратегию больших данных в своих собственных областях. Безусловно, другие государственные организации играют свою определённую роль в национальной стратегии больших данных. MPS и PLA, в частности, глубоко заинтересованы в больших данных для улучшения своей деятельности. Далее мы обратимся к этим организациям.

## **Глава третья**

### **Пример I. Предотвращение преступности и усиление внутреннего контроля**

Силы общественной безопасности Китая, пожалуй, с наибольшим энтузиазмом взяли на вооружение аналитику больших данных. В отсутствие правовых ограничений на использование правительством личных данных и в целях общественной безопасности такие мощные инструменты значительно расширяют возможности MPS по быстрому сопоставлению данных о судимости практически с любыми другими данными, которые имеют отношение к задержанию предполагаемых преступников<sup>18</sup>. Аналитика данных напрямую способствует

---

<sup>18</sup> "Китай: Полицейские системы "больших данных" нарушают конфиденциальность, нацелены против несогласия", Human Rights Watch, запись в блоге, 19 ноября 2017 года. Хотя Китайская Народная Республика недавно опубликовала относительно жесткие стандарты защиты частной жизни, они содержат значительные лазейки для свободного сбора и использования личных данных в целях национальной или общественной безопасности. С 1 октября 2018, стандарты были доступны от Национальной Стандартизации информационной безопасности Технический Комитет, «GB/T 35273-2017», 24 января, "[国家标准GB/T 35273-2017 《信息安全技术 个人信息安全规范》获批发布]", January 24, 2018.

строгую мониторингу и контролю граждан Китая, в том числе этнических меньшинств, и, что ещё более важно, выявлению любых признаков предполагаемой нелояльности или других нарушений. Эта возможность представляет большой интерес для ССР, поэтому мы должны ожидать постоянного внимания к тому, чтобы не только MPS, но и другие службы безопасности и разведки расширили использование этих новых инструментов. К концу 2020 года правительство планирует интегрировать все инструменты мониторинга и контроля на основе больших данных в так называемую национальную систему социального кредита – попытку присвоить репутационный рейтинг каждому человеку в китайском обществе. Будучи подключённым к сети, он будет представлять собой беспрецедентный и поистине оруэлловский уровень государственного наблюдения с помощью технических средств.

### Полицейское облако без границ

MPS использует новые наборы данных, которые он планирует централизовать в «полицейском облаке» (警务云)<sup>19</sup>. Облако полиции, которое в конечном итоге станет доступным для всех провинциальных и муниципальных органов полиции, повысит лёгкость, с которой полиция сможет устанавливать связи между разными базами данных, включая системы, не связанные с преступностью, такие как записи о жилье и занятости, для быстрого определения людей, мест и предприятий, представляющих интерес. Поскольку в Китае уже есть очень надёжная инфраструктура безопасности, основанная на технологиях, в частности, видеокамерах, следующим шагом служб безопасности является интеграция выходов этих устройств для обеспечения всеобъемлющего изображения наблюдения. Пекин уже регулярно сравнивает информацию из этих систем с национальными удостоверениями личности предполагаемых преступников. Хотя заставить все полицейские подразделения обмениваться информацией было непросто, MPS приводило удачные примеры провинций Шаньдун и Цзянсу, чтобы продемонстрировать ценность этих программ<sup>20</sup>.

В городе Вэйфан провинции Шаньдун полиция использовала повсеместную сеть камер видеонаблюдения для получения изображений, которые власти смогут задействовать технологией распознавания лиц для создания баз данных, способных

<sup>19</sup> “Public Security Police Cloud Infrastructure Construction and Design [公安警务云基础设施建设方案设计],” *Journal of People’s Public Security University of China (Science and Technology)* [中国人民公安大学警务信息工程学院], January 2016.

<sup>20</sup> See, for example, Wang Jincheng, “Technological Guidance to Service Actual Combat, Strongly Promoting the Construction and Application of Big Data Policing Cloud Computing Across the Entire Force [王金城——科技引导服务实战强力推进大数据警务云计算全警建设应用],” *Police Technology*, March 2015.

идентифицировать и отслеживать отдельных людей и транспортные средства<sup>21</sup>. Это позволило полиции находить и арестовывать подозреваемых после того, как их лица были обнаружены в общественных местах. Например, во время пивного фестиваля в Циндао полиция использовала распознавание лиц для выявления разыскиваемых подозреваемых в толпе, что привело к 25 арестам, а также использовала эту технологию для выявления и отказа в приёме наркозависимых<sup>22</sup>. Эта технология также позволила полиции проводить упреждающую работу путём сбора данных обо всех лицах данной категории, например на переполненном рынке; осуществлять отправку изображений в центр обработки данных полиции и связывать изображения с людьми, разыскиваемыми полицией<sup>23</sup>. Процесс работы полиции стал гораздо проще с программным обеспечением распознавания лиц. Согласно одному отчёту, полиция ежедневно получает текстовое сообщение, в котором сообщается, находится ли интересующее лицо в зоне их ответственности<sup>24</sup>. По отдельности, системы «мобильного опекуна» и «божественного глаза» Weifang могут автоматически идентифицировать любое транспортное средство по его номерному знаку, числу или другим характеристикам в течение секунд и отслеживать его движение. Полиция использовала эти системы для расследования 100% сообщений о краже мотоциклов и вернула все велосипеды их владельцам<sup>25</sup>. Кроме того, полиция использовала тенденции больших данных для оценки горячих точек преступности и соответствующего перенаправления активов<sup>26</sup>. Появляется ещё одна модель работы полиции из провинции Цзянсу, основанной на больших данных. Полиция города Суцзянь в провинции Цзянсу увеличила процент раскрытых дел на 85,1% благодаря своей работе по интеграции в свою систему больших данных из других полицейских органов. Они используют Географическую информационную систему полиции (PGIS) для стандартизации, интеграции и, где это возможно, геолокации 250 миллионов единиц данных. После побега подозреваемого в убийстве они поместили изображения лица подозреваемого с камер наблюдения в систему PGIS, которая смогла найти совпадения и идентифицировать его основные направления деятельности. В течение трёх часов полиция смогла найти и арестовать подозреваемого<sup>27</sup>. В более широком смысле полиция Цзянсу использовала большие данные для оптимизации распределения ресурсов в наиболее загруженных полицейских участках, тем самым повысив эффективность и снизив нагрузку на

<sup>21</sup> MPS, "Big Data Thinking Improves Police Innovation in Weifang, Shandong," September 22, 2015a; MPS, "Weifang, Shandong, Uses Big Data to Innovate a 3-D Law Enforcement System," October 7, 2015c. В 2015 году MPS призвал создать "вездесущую, полностью подключенную, всегда включенную и полностью управляемую" общенациональную сеть видеонаблюдения. See Josh Chin and Liza Lin, "China's All-Seeing Surveillance State Is Reading Its Citizens' Faces," *Straits Times*, July 8, 2017.

<sup>22</sup> "From Ale to Jail: Facial Recognition Catches Criminals at China Beer Festival," *The Guardian*, September 1, 2017.

<sup>23</sup> MPS, 2015c. Also see Paul Mozur, "Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras," *New York Times*, July 8, 2018.

<sup>24</sup> "China: Police 'Big Data' Systems Violate Privacy, Target Dissent," 2017.

<sup>25</sup> MPS, 2015a; MPS, 2015c.

<sup>26</sup> MPS, "Establishing the Big Data Public Security Concept [建立大数据公共安全理念]," September 28, 2015b.

<sup>27</sup> Zhao Jiabin [赵家新], Li Yirui [李沂瑞], and Tian Tingjiang [田廷江], "Shuyang, Jiangsu, Uses Big Data to Make Social Management Smarter [江苏宿迁运用大数据提升社会治理智能度]," MPS, November 28, 2016.

сотрудников этих участков<sup>28</sup>. Полиция также создала приложения для персональных устройств, которые позволяют людям более эффективно управлять взаимодействием с правоохранительными органами. Например, в Шуян, округе провинции Цзянсу, граждане взаимодействуют с полицией, платя штрафы и сообщая о преступлениях в Интернете. Полиция, в свою очередь, сообщает гражданским лицам такие важные новости, как обновлённую информацию о транспортном потоке<sup>29</sup>. В одном случае полицейское управление Суцзянь использовало свою систему PGIS для оптимизации перекрёстков и улучшения распределения дорожной полиции, уменьшив количество остановок движения на 39% и увеличив пропускную способность дорожного транспорта на 7,25%<sup>30</sup>. Такие виды взаимодействия могут укрепить доброжелательные отношения между людьми и правоохранительными органами. Действительно, по сообщениям, это или подобное приложение, которое также доступно в городе Ханчжоу, а также в Цзянсу и Суцзянь имеет рейтинг общественного одобрения 99,5 %<sup>31</sup>.

Помимо успешных примеров в Шаньдуне и Цзянсу китайские первоисточники указывают на множество других аспектов аналитики больших данных, позволяющих правоохранительным органам проводить операции или принимать более обоснованные решения гражданам. Например, полиция города Сямынь в провинции Фуцзянь раскрыла мошенническую схему с использованием почты после того, как заметила, что 26 человек, ранее обвинявшихся в мошенничестве с использованием почты, совместно снимали дом, были безработными в течение последних шести месяцев и недавно зарегистрировали кассовый автомат<sup>32</sup>. Город использовал аналогичную систему по интеграции различных битов данных для выявления рискованных и мошеннических платформ онлайн-кредитования<sup>33</sup>. В городской системе используется веб-сканер для сбора информации из СМИ, онлайн-источников и выявления аномалий, например, если имеются различия между официальным адресом и фактическим местом деятельности. Он также ищет жалобы, судебные иски, информацию о банковских счетах (наличных деньгах) и прибыльности. Он использует эти индикаторы для оценки финансовых платформ; проводит исследования платформ с плохими рейтингами. В будущем такую систему можно будет использовать для мониторинга деятельности частных компаний или фондовых бирж<sup>34</sup>. Некоторые другие муниципалитеты начали использовать большие

<sup>28</sup> Zhao, Li, and Tian, 2016.

<sup>29</sup> Zhao, Li, and Tian, 2016.

<sup>30</sup> Zhao, Li, and Tian, 2016.

<sup>31</sup> MPS, "Hangzhou, Jiangsu, Creates a Revolution in Public Security by Researching a Convenient App [浙江杭州警方研发便民服务APP用大数据助力公安改革]," May 2, 2016; Zhao, Li, and Tian, 2016.

<sup>32</sup> Zheng Liang [郑良], "Fujian Uses Big Data to Solve the People's Problems [福建大数据破解民生痛点]," MPS, January 14, 2017.

<sup>33</sup> MPS, "Online P2P Is Even Harder to Govern Than Internet Fraud! Xiamen Anti-Fraud Center Develops New Tool to Manage It [比电信诈骗更难治理的是网贷P2P!厦门市继反诈中心又出大数据监管利器]," January 12, 2017a.

<sup>34</sup> MPS, 2017a.

данные для отслеживания финансовых преступлений, отмывания денег и финансирования терроризма<sup>35</sup>.

Подобно тому, что сделала полиция в провинции Цзянсу, полицейское управление города Ухань в провинции Хубэй в партнёрстве с китайской компанией Alibaba разработало транспортное приложение, которое позволило сделать данные с камер дорожного движения общедоступными, уменьшая этим заторы на дорогах. Это было сделано после того, как транспорт стал одной из главных жалоб местных жителей в Интернете. В Шанхае водители без лицензии подвергаются высокому риску остановки из-за использования полицией технологии распознавания лиц. По состоянию на декабрь 2017 года около 835 водителей были пойманы и наказаны с момента начала общегородской кампании по регулированию дорожного движения в 2016 году. Эта же технология помогает полиции устранять неправильное вождение и даже более мелкие правонарушения, такие как пешеходный переход<sup>36</sup>.

## **В центре внимания этнические меньшинства**

Как отмечалось ранее, лидеры ССР особенно не доверяют некоторым этническим меньшинствам. Китайские службы безопасности используют все описанные выше методы полицейской деятельности на основе больших данных, а также дополнительные меры, нацеленные на эти группы населения, во имя сохранения внутренней стабильности. Самым крупным и, вероятно, наиболее агрессивным меньшинством являются уйгуры, этнически тюркская и религиозно мусульманская группа, члены которой проживают в основном в самой западной провинции Китая Синьцзян. Согласно докладу Хьюман Райтс Вотч от февраля 2018 г. о превентивной деятельности полиции в Синьцзяне, центральным элементом китайской слежки в провинции является Комплексная платформа совместных операций (IJOP) [体化联合作战平台]<sup>37</sup>. Хьюман Райтс Вотч отметила, что Бюро

<sup>35</sup> MPS, "Explanation of the State Council's Guidance Opinions on the Anti-Money Laundering, Anti-Terror Funding, Anti-Tax Evasion System [《国务院办公厅关于完善反洗钱、反恐怖融资、反逃税监管体制机制的意见》政策解读]," September 29, 2017b; MPS, "Press Conference on the State Council's Guidance Opinions on the Anti-Money Laundering, Anti-Terror Funding, Anti-Tax Evasion System [《国务院办公厅关于完善反洗钱、反恐怖融资、反逃税监管体制机制的意见》答记者问]," September 29, 2017c; MPS, "Publication of the State Council's Guidance Opinions on the Anti-Money Laundering, Anti-Terror Funding, Anti-Tax Evasion System [《国务院办公厅关于完善反洗钱、反恐怖融资、反逃税监管体制机制的意见》正式印发]," September 29, 2017d; State Council, "Internet Financial Risk Improvement Special Implementation Plan [互联网金融风险专项整治工作实施方案]," December 4, 2016a.

<sup>36</sup> "Facial Recognition Identifies Unlicensed Drivers in Shanghai," Xinhua, December 6, 2017.

<sup>37</sup> "China: Big Data Fuels Crackdown in Minority Region," Human Rights Watch, blog post, February 26, 2018. Источники, приведенные в этом разделе, взяты из анализа Human Rights Watch, но мы полностью реконструируем их здесь как переведенные цитаты для простых последующих исследований. Подробнее об IJOP см. "Секретарь Бюро общественной безопасности Лучжоу Ван Луцзюнь посетил уезд Ли", 17 мая 2017 года. Поскольку приобретение замечает объявление об учреждении IJOP, посмотрите, например, «Общественный Проект Контрольно-пропускного пункта безопасности Бюро безопасности Нетян Единственное Исходное Официальное уведомление Приобретения Перед завоеванием», 6 февраля 2017, и «Общественное Бюро безопасности графства Ечэна «Интегрированная Совместная Боевая

общественной безопасности Синьцзяна начало внедрять ИОР в августе 2016 года, и более ранняя программа завершена и готова к эксплуатации в префектуре Кашгар на юго-западе Синьцзяна. ИОР использует распознавание лиц, инфракрасные камеры и анализаторы Wi-Fi, которые обнаруживают идентифицирующие адреса компьютеров и других персональных устройств для анализа собранных личных данных<sup>38</sup>. Эти датчики размещаются в местах, представляющих большой интерес, например, за пределами домов религиозных деятелей или рядом с развлекательными заведениями. Хьюман Райтс Вотч также сообщала, что китайские власти дополняют эти наборы данных, собирая генетическую информацию, отпечатки пальцев, сканирование радужной оболочки глаза и группы крови этнических меньшинств<sup>39</sup>.

Контрольно-пропускные пункты физической безопасности и системы управления посетителями – системы, которые проверяют всех посетителей перед входом в сообщества с ограниченным доступом (сообщества, которые позволяют входить только утверждённым людям) – предоставляют дополнительные данные о местонахождении и действиях целей<sup>40</sup>. Представители службы безопасности Китая дополняют эти данные наблюдений с помощью регулярных личных интервью для определения «идеологической ситуации» их целей и иного расследования того, что власти считают имеющим отношение к поддержанию внутренней стабильности<sup>41</sup>. Предиктивная полицейская деятельность находится в центре внимания Китайского Народного университета общественной безопасности, China Electronics Technology Group Corporation, и Специального следственного подразделения Бюро общественной безопасности Синьцзяна. Хьюман Райтс Вотч обнаружила, что список полиции, содержащий не менее 75 поведенческих индикаторов, теперь используется в ИОР, что позволяет сотрудникам службы безопасности не только проводить анализ больших данных, но и корректировать алгоритмы, используя

---

Платформа» Второй Проект оборудования Фазы Единственная Исходная Реклама Приобретения KSYCX (DY) 2017-29. [叶城县公安局“一体化联合作战平台”二期设备项目单一来源采购公示 KSYCX (DY) 2017-29号], August 24, 2017.

<sup>38</sup> See, for example, “Ebin County, Minta County, ‘Integrated Joint Combat Platform’ Video Surveillance Construction Project Publicity Results [塔城区额敏县“一体化联合作战平台”视频监控建设项目成交结果公示],” March 21, 2017. Другие источники, которые приводит Human Rights Watch - например, по распознаванию лиц и Wi-Fi-нюхателям, - были удалены по состоянию на 8 мая 2018.

<sup>39</sup> “China: Minority Region Collects DNA from Millions: Private Information Gathered by Police, Under Guise of Public Health Program,” Human Rights Watch, blog post, December 13, 2017.

<sup>40</sup> Источники Human Rights Watch по этим вопросам были удалены, а в случае с системами управления посетителями на сайте системы было написано “информация, которую вы хотите просмотреть, не существует или еще не утверждена!”. Страница была перемещена или удалена.” [您要查看的信息不存在或者还未通过审批! ]

<sup>41</sup> “China: Big Data Fuels Crackdown in Minority Region”, 2018. Очные собеседования могут проводиться каждый день командой под названием Fang Huiju, чья работа заключается в проведении собеседований. Смотрите, например, “2017” Fang Huiju “Visit Newsletter Issue 5 20160301”, [2017 “05-20160301], March 1, 2016. Китайские службы безопасности проверяют на предмет “необычности”. См., например, Hetian Town Government, “Hetian Area Emergency Management Work Information Phase 1”, January 6, 2017. Оценка Human Rights Watch содержит цитату из заявления о том, что ИОР рассматривает все, что “связано со стабильностью”. Однако с тех пор заявление Китая по этому вопросу было исключено. See, for example, Hetian Town Government, “Hetian Area Emergency Management Work Information Phase 1 [和田地区应急管理工作信息 第1期],” January 6, 2017. The Human Rights Watch assessment has a citation for the statement that ИОР looks into anything “related to stability.” However, the Chinese statement to that effect has since been deleted.

отзывы ИОР<sup>42</sup>. Лица, которые демонстрируют факторы риска, такие как зарубежные поездки большой продолжительности, проактивную поддержку местных мечетей, аномальное потребление электроэнергии, отношения с другими людьми, у которых есть факторы риска или использование приложений, помогающих избежать наблюдения, автоматически отмечаются системой. Эти люди могут быть подвергнуты полицейскому расследованию, внесудебному задержанию или ограничению посещения общественных мест<sup>43</sup>.

Документы, просочившиеся в *New York Times* в конце 2019 года, предполагают, что ССР придерживается почти эпидемиологического взгляда, сравнивая экстремизм с вирусом, который должен быть устранён путём строгого бессрочного обязательного перевоспитания, что привело к лишению свободы тысяч уйгуров и других людей, мусульман в Синьцзяне. Документы также предполагают, что поведение как заключённых, так и их семей будет продолжаться оцениваться для получения балльной оценки, которая, в свою очередь, будет определять продолжительность их заключения<sup>44</sup>. Неясно, как ИОР или другие методы, основанные на данных, опробованных в Синьцзяне, впишутся в более широкую инфраструктуру облака полиции, хотя не исключено, что многие функции ИОР в конечном итоге будут реализованы полицией за пределами Синьцзяна<sup>45</sup>.

## **Национальная система социального кредитования: скоро появится**

В 2014 году Государственный совет объявил, что Китай добивается создания национальной системы социального кредитования.<sup>46</sup> Как упоминалось ранее, национальная система социального кредита – это попытка правительства присвоить репутационный рейтинг каждому человеку в китайском обществе. Рейтинги можно использовать для предотвращения доступа людей с низким рейтингом к определенным социальным благам и для поощрения высокопоставленных лиц улучшенным доступом к таким благам. Эта система, которую предлагается запустить в 2020 году, будет использовать методы анализа больших данных, уже используемые различными китайскими компаниями и муниципалитетами.

<sup>42</sup> “Identify 75 Religious Extreme Activities [识别75种宗教极端活动],” *Nanchang Gonganjū* [南昌公安局], September 8, 2015. For more on Chinese conceptual thinking on algorithms, see “An Anti-Terrorism Risk Analysis Method Based on Bayesian Theory [一种基于贝叶斯理论的反恐风险分析方法],” *China National Knowledge Infrastructure*, September 2017.

<sup>43</sup> “China’s Algorithms of Repression: Reverse-Engineering a Xinjiang Police Mass Surveillance App,” Human Rights Watch, May 1, 2019.

<sup>44</sup> Austin Ramzy and Chris Buckley, “‘Absolutely No Mercy’: Leaked Files Expose How China Organized Mass Detentions of Muslims,” *New York Times*, November 16, 2019.

<sup>45</sup> “China’s Algorithms of Repression: Reverse-Engineering a Xinjiang Police Mass Surveillance App,” 2019.

<sup>46</sup> “Planning Outline for the Construction of a Social Credit System (2014–2020),” *China Copyright and Media*, blog post, June 14, 2014, updated April 25, 2015. Also see State Council, “State Council Guidance Principles for the Building of the Social Credit System [国务院办公厅关于加强个人诚信体系建设的指导意见],” December 30, 2016b.

Проект начался, когда Пекин выдал лицензии восьми компаниям на тестирование их алгоритмов для того, чтобы выяснить, подходят ли эти возможности для использования правительством как для защиты своих граждан, так и для их контроля. Сообщается, что Пекин в значительной степени полагается на Tencent (разработчик популярной многоцелевой платформы обмена сообщениями WeChat) и Alibaba (китайскую службу онлайн-покупок и доставки, аналогичную Amazon) в качестве доказательств для создания цифровых подключений между огромными и разрозненными наборами данных<sup>47</sup>. В частности, Группа оценки кредитной истории, управляемая Alibaba, известная как Sesame Credit, была в центре внимания по планированию системы социального кредита. В отличие от США, где кредитная история оценивается в вакууме, Sesame Credit и другие подобные компании комбинируют проверки кредитной истории с другими показателями «хорошего» или «плохого» поведения. Это не должно вызывать удивления, учитывая, что ССР с самых первых дней своего существования вела досье, известные как данган, на всех людей, чтобы оценить надёжность человека и определить, требуется ли для него политическое перевоспитание или другая форма наказания. С этой точки зрения национальная система социального кредитования будет, по сути, оцифрованной версией досье ССР 21 века о населении.

Sesame Credit продемонстрировал методы преобразования личных данных в рейтинги надёжности, которые, вероятно, будут, по крайней мере, частично включены в обязательную национальную систему социального кредитования в 2020 году. Новая система будет работать за счёт объединения проверок кредитной истории с проверками многих десятков, а может быть, сотен других баз данных о деятельности граждан и их образе жизни<sup>48</sup>. В Sesame Credit от Alibaba эти базы данных организованы по четырём ключевым областям, помимо личной кредитной истории. Во-первых, существует «способность пользователя выполнять свои договорные обязательства», предполагая, что это мера того, достаточны ли данные о статусе занятости и заработной платы для выполнения своих обязанностей<sup>49</sup>. Во-вторых, все люди должны проверить свою личную информацию и характеристики через официальные запросы. В-третьих, выбор образа жизни индивида представлен наборами данных об их покупательском поведении, где и как индивид проводит свое время, и что он говорит в Интернете и в других местах, и эти результаты тщательно изучаются. И, наконец, оцениваются сотрудники человека, оказывающие на него соответствующее влияние.

Как только система перейдёт в режим онлайн, оценки граждан будут обнародованы, чтобы убедить людей избежать стыда за счёт улучшения

---

<sup>47</sup> Rachel Botsman, "Big Data Meets Big Brother as China Moves to Rate Its Citizens," *Wired UK*, October 21, 2017.

<sup>48</sup> По крайней мере 30 местных органов власти уже поддерживают эту инициативу. See "China Invents the Digital Totalitarian State," *The Economist*, December 17, 2016.

<sup>49</sup> Alibaba Group, "Ant Financial Unveils China's First Credit-Scoring System Using Online Data," press release, Hangzhou, China, January 28, 2015.

результатов. Ассоциация оценок даёт очевидный стимул держаться подальше от «плохих» людей. Похоже, что это намеренное нацеливание на разделение групп и, следовательно, на предотвращение возникновения беспорядков, которые могут угрожать стабильности режима ССР. Например, в рамках национальной системы социального кредитования политический диссидент, который публикует негативные отзывы о правительстве в Интернете, скорее всего, окажется изолированным от общества, потому что теоретически никто не захочет понизить свою оценку из-за связи с диссидентом с низким рейтингом<sup>50</sup>. Будьте уверены, что политическая сфера – лишь одна из многих сфер поведения, которые могут повлиять на оценку человека. Индивидуальные покупки будут тщательно изучаться, чтобы определить трудовую этику человека. Того, кто покупает подгузники, можно считать более ответственным, чем того, кто покупает видеоигры, потому что последнее может указывать на молодость и лень<sup>51</sup>. В конечном счёте, ССР стремится измерить уровень доверия каждого гражданина, как это определено намерением. ССР может стимулировать «хорошие» оценки, предлагая более лёгкий доступ к кредитам, аренде без залога, ускоренные визы и льготный доступ к сайтам знакомств. Предлагаемые штрафы за более низкие баллы включают ограниченный доступ к образованию и жилью, трудности с получением государственного финансирования, а также ограничения на поездки<sup>52</sup>.

Идея наличия «системы вознаграждений» не нова в коммунистическом Китае. Каждый год Пекин награждает сотни тысяч людей всевозможными наградами, такими как «выдающиеся кадры» или «цивилизованная деревня», призами, варьирующимися от денег до более высокой пенсии, лучшего медицинского страхования и государственного жилья<sup>53</sup>. Оцифровка этой системы с масштабным размахом подкрепляет ожидания руководства ССР о том, что к концу 2020 года режим сможет править страной, которая действительно «позволит [нам] надёжным бродить повсюду под небесами, в то же время затрудняя для дискредитированных возможность сделать хотя бы один шаг»<sup>54</sup>.

Следует отметить, что этот проект не является бесспорным в Китае. Летом 2019 года правительство округа Вулянь в провинции Шаньдун подверглось критике за внесение учителя в чёрный список социального кредита округа за применение телесных наказаний к учащимся, при этом некоторые пользователи сети утверждали, что кредитоспособность должна быть исключительно финансовым вопросом. Хотя школа учителя приняла административные меры против него (относительно бесспорного последствия применения им телесных наказаний), в

---

<sup>50</sup> Rick Falkvinge, "In China, Your Credit Score Is Now Affected by Your Political Opinions— and Your Friends' Political Opinions," *Privacy News Online*, blog post, October 3, 2015.

<sup>51</sup> Botsman, 2017.

<sup>52</sup> "China Invents the Digital Totalitarian State," 2016.

<sup>53</sup> "China Invents the Digital Totalitarian State," 2016.

<sup>54</sup> Amy Hawkins, "Chinese Citizens Want Their Government to Rank Them," *Foreign Policy*, May 24, 2017.

итоге он был исключён из чёрного списка, а статья с критикой решения внести его в чёрный список была перепостена на сайте «Peoples Daily», поддерживаемой ССР, указывая на определённый уровень официальной поддержки или, по крайней мере, терпимости к его взглядам<sup>55</sup>. Точно так же в начале 2018 года граждане Китая с возмущением отреагировали, когда китайская компания Ant Financial начала зачислять их в свою систему социального кредитования без их ведома<sup>56</sup>. По состоянию на июль 2020 года, хотя некоторые компании и местные органы власти продолжают экспериментировать с системами социального кредитования, остаётся неясным, как Пекин будет формировать свою национальную политику социального кредитования и в какой степени общественное недовольство будет удерживать его от реализации в национальном масштабе некоторых из наиболее драконовских мер, используемых в Синьцзяне.

### **COVID-19: Генеральная репетиция контроля над населением?**

В последние дни 2019 года и в 2020 году болезнь, вызванная уникальным штаммом коронавируса, COVID-19, бушевала по всему Китаю. Как только Пекин решил вмешаться и рассматривать эпидемию как чрезвычайную ситуацию в стране, в дело вмешались китайские технологические, коммуникационные и транспортные компании. Телекоммуникационные компании (предположительно, используя свой доступ к данным о местоположении, генерируемым смартфонами) и туристические компании начали отслеживать людей, которые путешествовали в места с высоким уровнем заражения или из этих мест, часто публикуя списки этих людей вместе с указанием способа передвижения, который они предприняли, чтобы любой из тех, кто с ними контактировал, мог быть протестирован или помещён в карантин<sup>57</sup>. Китайские телекоммуникационные компании разработали услугу, позволяющую пользователям получать отчёт о своём местонахождении за последние 14 дней, чтобы люди могли доказать работодателям, квартирным комитетам или другим лицам, что они не были в районе, где произошла серьёзная вспышка болезни<sup>58</sup>. И Tencent, и Alibaba разработали приложения, помогающие людям узнать, когда они могут подвергнуться риску заражения вирусом<sup>59</sup>. Приложение Alipay Health Code от Alibaba, которое, как сообщается, использует алгоритм искусственного интеллекта

<sup>55</sup> Gao Lu [高路], "Credit Blacklists: Not the Solution to Every Social Problem [信用黑名单, 不能啥都往里装]," *Qianjiang Evening News* [钱江晚报], trans. Ryan Soh and Jeffrey Ding for *ChinAI*, No. 61, July 30, 2019.

<sup>56</sup> Paul Mozur, "Internet Users in China Expect to Be Tracked. Now, They Want Privacy," *New York Times*, January 4, 2018. Note that many Chinese users continue to use Ant Financial's apps.

<sup>57</sup> Hu Yong, "The Public Interest and Personal Privacy in a Time of Crisis (Part I)" *Weixin*, trans. Jeff Ding for *ChinAI*, No. 85, March 6, 2020a.

<sup>58</sup> Shawn Yuan, "How China Is Using AI and Big Data to Fight the Coronavirus," *Al Jazeera*, March 1, 2020.

<sup>59</sup> Pratik Jakhar, "Coronavirus: China's Tech Fights Back," *BBC Monitoring*, March 3, 2020.

для анализа сообщаемых симптомов, истории поездок и данных о местоположении, чтобы классифицировать пользователей как зелёных, жёлтых или красных в зависимости от риска заражения этим заболеванием, было принято как минимум 100 муниципалитетами, и имеются сообщения о работодателях, управляющих квартирами и даже о сотрудниках полиции на съездах с автострад, отказывающихся в доступе любому, у кого нет «зелёного» балла<sup>60</sup>.

Как и в случае с системой социального кредитования, усилия Китая по контролю над населением во время пандемии не обошлись без разногласий. Ху Юн, профессор Школы журналистики и коммуникации Пекинского университета, чей блог насчитывает 800 000 подписчиков, раскритиковал ответ Китая за то, что он выдвигает больше требований к правам личности, чем это необходимо, даже с учётом кризиса<sup>61</sup>. Многие китайские граждане из Хубэй (провинция в эпицентре пандемии) жаловались на широко распространённую дискриминацию, которой во многом способствовало официальное разглашение личной информации, национальные удостоверения личности, которые сразу же могут определить местонахождение официального места жительства гражданина, и, возможно, даже определение непрозрачным путём в соответствии с приложением Кодекса здравоохранения Alipay<sup>62</sup> риска человека и его возможности вернуться домой или на работу, что вызвало возмущение некоторых пользователей<sup>63</sup>.

Следует отметить, что перед лицом этой вспышки заболеваемости Китай является не единственной страной, ограничивающей некоторые личные свободы для предотвращения заражения, и во всем мире такие шаги часто поддерживаются местным населением<sup>64</sup>.

Что потенциально беспокоит в мерах Китая, так это то, что инструменты, которые ССР разрабатывает во время кризиса для отслеживания своих граждан и ограничения их автономного передвижения с использованием личных данных могут быть позже применены в более репрессивных целях. Как отмечалось в предыдущих разделах, китайское правительство и китайские компании (а также другие компании и правительства по всему миру) давно собирают данные о поездках, местонахождении, здоровье людей и множестве других видов деятельности. Китай, однако, может быть уникальным по масштабу готовности заставить работодателей, телекоммуникационные компании, полицию, больницы, предприятия и местные органы власти объединять свои данные, а затем автоматически вводить ограничения для граждан в результате этих сведений. Как заметил Ху Юн: «Но вы также можете

---

<sup>60</sup> Paul Mozur, Raymong Zhong, and Aaron Krolik, "In Coronavirus Fight, China Gives Citizens a Color Code, with Red Flags," *New York Times*, March 1, 2020.

<sup>61</sup> Hu Yong, "The Public Interest and Personal Privacy in a Time of Crisis (Part II)" *Weixin*, trans. Jeff Ding for *ChinAI*, No. 86, March 6, 2020b.

<sup>62</sup> Hu Yong, 2020a; Hu Yong, 2020b; Jakhar, 2020.

<sup>63</sup> Mozur, Zhong, and Krolik, 2020.

<sup>64</sup> Grace Moon, "This Is How South Korea Flattened Its Coronavirus Curve," *NBC News*, March 24, 2020.

спросить себя: знает ли история пример того, что когда у правительства есть инструменты слежки, оно будет сохранять скромность и осторожность при их использовании?»<sup>65</sup>.

## Всеведение с китайскими особенностями

В этом разделе показано, что цель ССР – использовать возможности анализа больших данных для строгого и всестороннего мониторинга и контроля собственного населения. Китай буквально создаёт систему наблюдения и больших данных на национальном уровне, которая конкурирует даже с самыми смелыми вымышленными отчётами, описанными в известной антиутопической литературе, например, Джорджа Оруэлла в 1984 году. Аналогично рейтинговой системе, изображенной в эпизоде *Black Mirror* «Порыв» (серия 3, эпизод 1, октябрь 2016 г.), оценка социального кредита теоретически будет ограничивать индивидуальные действия, поскольку граждане будут стремиться к достижению преимуществ высоких оценок и избегать опасностей низких оценок<sup>66</sup>. Некоторые из наиболее амбициозных проектов ССР (особенно в Синьцзяне), в некотором роде, также похожи на технологию «precrime», (предварительное преступление) изображённую в фильме «Minority Report» 2002 года (на основе одноименного романа Филипа К. Дика 1956 года), в котором полиция может прогнозировать и предотвращать преступления прежде, чем они произойдут. ССР, кажется, считает, что многие преступления можно предотвратить, собирая как можно больше данных об отдельных гражданах, а затем осуществляя превентивное наблюдение или даже арестовывая тех, кто представляет определённые факторы риска. Пекин надеется, что такие методы позволят ему остановить преступность или политическое инакомыслие, так что прежде, чем это когда-либо произойдёт «предварительное восстание» – здесь это будет точным описанием, самых непослушных граждан превратят в более стоворчивых людей. Пекин уже использует эти методы и добавляет новое коварное измерение, стимулирующее хорошее поведение через кредитную систему. Более того, Пекин, похоже, верит в эти возможности для определения областей, особенно подверженных действиям, вызывающим беспокойство, и, соответственно, перенаправляет ресурсы на решение этих проблем. В конечном итоге китайские лидеры хотят создать способность прогнозирования полицейской деятельности для достижения более полной способности контролировать общественное взаимодействие.

<sup>65</sup> Hu Yong, 2020b.

<sup>66</sup> Mara Hvistendahl, “Inside China’s Vast New Experiment in Social Ranking,” *Wired*, December 14, 2017. Хотя ни одно крупное СМИ, насколько мы могли обнаружить, не установило связи между антиутопическим взглядом *Black Mirror* и системой социального кредита, некоторые блогеры сделали это, хотя есть разногласия относительно того, насколько точно сравнение с китайской системой. See, for example, Mianzi Jingxuan, “Are They Approaching Hell? China Will Implement Black Mirror’s Scoring System [面子精选], [他人即地狱? 中国要实行《黑镜》里的评分制了],” *Sohu*, December 14, 2017.

## ГЛАВА ЧЕТВЕРТАЯ

### Практический пример II: Повышение боевых возможностей Народно-освободительной армии

Во время выступления на 19-м съезде партии в октябре 2017 года председатель Си озвучил своё намерение, касающееся полной трансформации PLA в «силы мирового класса» к 2050 году<sup>67</sup>. В поддержку этой долгосрочной цели Китай считает аналитику больших данных жизненно важным национальным ресурсом. Пекин проявил особый интерес к использованию больших данных – и, в конечном итоге, искусственного интеллекта – для улучшения широкого спектра возможностей PLA, которые мы подробно описываем в этой главе. Общая идея китайских первоисточников заключается в том, что владение аналитикой больших данных будет способствовать победам Китая в будущих военных конфликтах между великими державами.

### Определение больших данных для Народно-освободительной армии

Хотя общедоступного всеобъемлющего документа о планах PLA для больших данных не существует, обзор множества общедоступных материалов позволяет акцентировать внимание на большие данные в соответствии с прерогативами национальной обороны. В одной статье, опубликованной в июне 2016 года, в общих чертах описываются большие данные национальной обороны (国防大数据)<sup>68</sup>. Авторы статьи определили большие данные национальной обороны как коллективные данные, генерируемые такими военными действиями, как защита национального суверенитета, единства, территориальной целостности и безопасности и ресурсы данных, генерируемые политической, экономической, научной, дипломатической, образовательной и другой деятельностью, связанной с военными проблемами.

Ключевые характеристики больших данных национальной обороны выходят за рамки общих больших данных (высокая скорость, разнообразие, большой объём и высокая ценность) и включают в себя характеристики 6S, включая сверхсложность

<sup>67</sup> “Full Text of Xi Jinping’s Report at the 19th CPC National Congress,” 2017.

<sup>68</sup> He You [何友], Zhu Yangyong [朱扬勇], Zhao Peng [赵鹏], Chai Yong [柴勇], Liao Zhicheng [廖志成], Zhou Wei [周伟], Zhou Xiangdong [周向东], Wang Haipeng [王海鹏], Wang Wei [汪卫], Xiong Yun [熊贇], Xu Zhoujun [许舟军], Peng Xuan [彭煊], Meng Hui [孟晖], and Wang Shengjin [王生进], “Panorama of National Defense Big Data [系统工程与电子技术],” *Systems Engineering and Electronics* [国防大数据概论], Vol. 6, 2016.

[复杂性], сверхсекретность [超保密性], быстрое развёртывание [高机动性], безопасность [高安全性], сильную степень конфронтации [强对抗性] и строгую своевременность [强实时性]. Сверхсложность означает, что военные данные во многих случаях труднее обрабатывать, чем общие данные. Сверхсекретность подчёркивает важность защиты больших данных национальной обороны от кражи или нападения враждебных сил. Быстрое развёртывание определяется как возможность гибкого использования военных данных в нескольких областях ведения войны по запросу. Безопасность описывает необходимость защиты инфраструктуры, поддерживающей анализ больших данных, например электроэнергии и компьютерного оборудования. Сильная степень конфронтации указывает на то, что создание потенциала больших данных для национальной обороны является непростой конкурентной задачей, реализующейся в контексте отказа от сотрудничества. Наконец, высокая своевременность описывает операционную среду, в которой решение должно приниматься быстро из-за постоянно меняющейся ситуации на поле боя.

### **Большие данные национальной обороны – ключ к победе в борьбе великих держав**

Развивая эти характеристики 6S, Китай стремится укрепить свою главную цель по созданию вооружённых сил, способных «побеждать в локальных информационных войнах» [打赢信息化局部战争]<sup>69</sup>. С тех пор, как эта концепция впервые появилась в 2015 году в самом последнем техническом документе Китая по вопросам обороны, она получила широкую интерпретацию как развёртывание трансформированной PLA, способной контролировать – и однажды доминировать – в среде информационных войн, чтобы успешно выигрывать войны в своём собственном регионе, прежде всего на Тайване, а также в Восточно-Китайском и Южно-Китайском морях. Действительно, согласно многочисленным китайским источникам, умение использовать большие данные для национальной обороны стало ключом к победе в войнах такого типа против других передовых государств.

Например, по словам Ли Дагуанга из Департамента военной логистики и военного научно-технического оборудования при Национальном университете обороны Китая, большие данные национальной обороны стали новым фокусом в

---

<sup>69</sup> *Победа в информационных локальных войнах фактически является девятым пересмотром китайских стратегических руководящих принципов с момента основания Китайской Народной Республики. Дополнительные сведения об этой истории см. в разделе [打赢信息化局部战争] [军事战略方针], see M. Taylor Fravel, "China's New Military Strategy: 'Winning Informationized Local Wars,'" *China Brief*, Vol. 15, No. 13, July 2, 2015.*

борьбе между великими державами<sup>70</sup>. Ли далее высказал мнение, что суверенитет данных в сетевом пространстве так же важен, как суверенитет земли, моря, воздуха и космоса. Большие данные станут важным компонентом противоборства систем, которое, по оценке PLA, станет главной темой войны в 21 веке. По словам Шан Целиана, директора Департамента мобилизации в Нанкинском армейском командном училище, системная конфронтация и совместные операции производят большой объём данных. Только всесторонне, точно и быстро выявляя ключевые моменты из сложных источников данных, мы можем эффективно организовать противостояние, прорваться сквозь туман войны и достичь точности в плане ведения войны<sup>71</sup>.

Лю Линьшань, исследователь Центра информационных исследований Академии военных наук PLA, отметил, что с наступлением «эры больших данных» большие данные национальной обороны теперь играют решающую роль в качестве «стратегического ресурса»<sup>72</sup>. Распространение больших данных сигнализирует о появлении новой формы войны, в основе которой лежит как нарушение, так и защита данных. Победа будет достигнута путем разграбления и уничтожения ресурсов данных противника, а также создания превосходства и использования данных для быстрого принятия оперативных решений, ведущих к успеху. По словам Лю, данные особенно важны для проведения совместных операций. Только с эффективным потоком данных посредством их обмена, интеграции и взаимодействием между различными оперативными элементами можно преодолеть когнитивные ограничения командиров, что сокращает время, необходимое для принятия точных решений. Большие данные также способствуют более эффективному управлению военными ресурсами.

Мы также ознакомились с китайскими работами национальной обороны по большим данным на оперативном уровне. В подтверждение видения Лю командующий Северным театральным командованием Китая генерал-лейтенант Ли Цяомин, например, определил, что принятие решений командования должно стать автономным за счёт беспрепятственного использования данных, полученных от различных подразделений, служб и мест ведения боевых действий<sup>73</sup>. По словам Ли, PLA должна использовать технологии с использованием облачных вычислений и понятий Интернета. Плоские структуры, простые и интегрированные иерархии и системная интеграция представляют собой текущее направление развития систем оперативного управления, которые в конечном итоге создадут почву для

<sup>70</sup> Cheng Rong [程荣] and Pei Xian [裴贤], "National Defense Big Data: The Intelligent Core That Protects National Security [国防大数据：守护国家安全的智慧芯] ," *People's Liberation Army Daily* [解放军日报], January 11, 2017.

<sup>71</sup> Cheng and Pei, 2017.

<sup>72</sup> Liu Linshan [刘林山], "What Is the Function of Big Data in the Construction of a Modern National Defense and the Military? [大数据在国防和军队现代化建设中有何作用? ]," *People's Liberation Army Daily*, February 1, 2018.

<sup>73</sup> Li Qiaoming [李桥铭], "Big Data: Making War Commanding Decisions More Scientific [大数据：让战争指挥决策更科学]," *People's Liberation Army Daily*, February 28, 2017.

«интеллектуальности» [智能化] войны. Для Пекина явное использование термина «интеллектуальность», будь то в военной или гражданской сфере, указывает на более высокий уровень автоматизации. Хотя большинство учёных из PLA придерживаются мнения, что ИИ никогда полностью не заменит человеческих оперативных командиров, они все же верят, что он может действовать как «цифровой офицер штаба», способный собирать и представлять разведанные о противнике, выявлять намерения противника и контролировать операции<sup>74</sup>.

Одно из наиболее обширных описаний значения больших данных для PLA принадлежит заместителю командира Центрального командования театра военных действий генерал-лейтенанту Чжан Сюйдун<sup>75</sup>. Будучи командующим сухопутными войсками Центрального театра военных действий, Чжан обратил внимание на четыре основных момента. Во-первых, полевые командиры должны чётко осознавать новые изменения, продиктованные большими данными. Информационное превосходство стало новым механизмом победы в бою, а принятие решений на основе данных стало новой парадигмой принятия боевых решений. Во-вторых, командиры должны сохранять ясность в отношении общей стратегии построения больших военных данных. Они должны стремиться к взаимодействию гражданских и военных усилий, расширять каналы ресурсов больших данных, строить систему хранения больших данных и повышать качество и эффективность потока и передачи больших данных. В-третьих, командиры должны использовать боевой потенциал больших военных данных. Это включает в себя использование преимуществ больших данных путём сосредоточения внимания на противнике и соблюдения требований линии фронта. Это также означает исследования и разработки программного обеспечения, специализирующегося на больших данных в военной сфере, создание мощных платформ боевого управления большими данными и создание сил боевой поддержки больших данных. Наконец, командиры должны поддержать инициативу использования военными больших данных. Чжан считал, что это повлечёт за собой перехват инициативы в стремлении стать экспертами в области использования больших данных, эксплуатации программного обеспечения для больших данных и функций планирования больших данных.

---

<sup>74</sup> Yuan Yi [袁艺], "Will AI Command Future Wars? [人工智能将指挥未来战争?]," *Defense Daily* [中国国防报], January 12, 2017.

<sup>75</sup> Zhang Xudong [张旭东], "Zhang Xudong: Seize the Initiative in Future Battlefield Data [张旭东: 抢占未来战场数据主导先机]," *People's Liberation Army Daily*, August 31, 2017.

## Военное применение больших данных

Эти типы концептуализации побудили PLA изучить потенциальные возможности применения больших данных для национальной обороны. Основные сферы интересов включают в себя управление, контроль, связь, компьютеры, наблюдение и разведку (C4ISR); оборудование и обслуживание; логистику; здравоохранение; мобилизацию; обучение; набор персонала; моделирование и симуляцию; кибербезопасность.

## Базовая инфраструктура больших данных

Прежде чем PLA сможет использовать аналитику больших данных для увеличения боевого потенциала, он должен сначала создать необходимую инфраструктуру больших данных. Поэтому PLA планирует значительно усилить свои возможности по хранению, передаче и анализу больших данных национальной обороны<sup>76</sup>. Пекин должен создать массивную и «контейнерную» платформу больших данных национальной обороны, предполагая, что система должна иметь чётко определённые границы инфраструктуры и, возможно, конкретные наборы миссий. Об этом говорится в статье, подготовленной совместно Институтом смешанных информационных исследований военно-морского авиационного инженерного колледжа, кафедрой электротехники Университета Цинхуа, Фуданьским университетом, Колледжем компьютерных наук, Пекинским научно-исследовательским институтом дистанционного зондирования и Пекинским научно-исследовательским институтом информационных технологий<sup>77</sup>. Авторы этой статьи выступили за дальнейшую «встроенную» концепцию национальных оборонных систем больших данных, которые позволят PLA одержать верх в будущем противостоянии интеллектов. В отдельном тендерном предложении PLA объявила о своём намерении приобрести системы хранения данных с централизованным развёртыванием, управлением, мониторингом и безопасностью. Каждая из этих систем будет включать в себя два узла управления, узлы данных, сервисные коммутаторы, коммутаторы управления (как в телекоммуникациях) и всё необходимое программное обеспечение. Каждый узел данных будет обладать 16 гигабайтами памяти, а система – 40 терабайтами памяти<sup>78</sup>. В другой статье Китайской академии электроники и информационных технологий утверждается, что

<sup>76</sup> Han Ming [韩明], Yang Jibao [杨继宝], and Lu Xiang [卢祥], “The Era of Big Data War [大数据战时代],” *China New Telecommunications* [中国心通信], Vol. 11, 2017.

<sup>77</sup> He You et al., 2016.

<sup>78</sup> “Big Data Machine [大数据一体机],” *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], April 5, 2016.

система командной разведки должна иметь возможность интегрировать множество различных типов данных, как с мест ведения боевых действий, так и общедоступных. Такая система может использовать архитектуру и инструменты Hadoop, такие как Hadoop Storm и Hadoop Spark, предназначенных для сбора, хранения и обработки больших объёмов данных. Она будет обслуживать самых разных военных заказчиков, от объединённых командных центров до отдельных солдат<sup>79</sup>. Следует отметить, что Академия раннего предупреждения ВВС PLA (PLAAF) также проводит работу по совершенствованию методов планирования Hadoop для анализа разведывательных данных и Народная вооружённая полиция, по-видимому, заинтересована в этом<sup>80</sup>. Однако Пекин признаёт, что одного построения инфраструктуры больших данных национальной обороны будет недостаточно, поскольку для того, чтобы управлять этими системами и их эксплуатировать, требуется человеческий талант. Например, по данным Лаборатории моделирования сложных электрических систем при Колледже оборудования, многие люди говорят о моделировании и больших данных, но на самом деле не понимают механизмов, которые их связывают. Таким образом, Колледж оборудования стремился изложить две «мостовые» теории, предположительно, для их соединения<sup>81</sup>. Между тем, в редакционной статье говорилось, что PLA должна работать над совершенствованием своих правил больших данных и нанимать больше людей, имеющих опыт работы с большими данными<sup>82</sup>.

## **Командование, управление, связь, компьютеры, разведка, наблюдение и разведка**

PLA в целом, и особенно ВМС PLA (PLAN) и PLAAF, явно заинтересованы как в использовании национальной обороной больших данных для интеграции выходных данных с нескольких датчиков разных типов, так и в использовании глубокого обучения для выявления распознаваемых характеристик целей для улучшения и автоматизации их дифференциации. Такая интеграция данных должна

<sup>79</sup> Guo Jiguang [郭继光] and Huang Sheng [黄胜], "Study on Architecture of Big Data Based on Military Intelligence Analysis and Service System [基于大数据的军事情报分析与服务系统架构研究]," *Journal of the Chinese Academy of Electronics and Information Technology* [中国电子科学研究院学报], Vol. 4, 2017.

<sup>80</sup> Jiang Surong [蒋苏蓉] and Meng Jiangqiao [蓝江桥], "Timeout Prediction of the Schedule Method for Big Data of the Intelligence Analysis Based on Hadoop [Hadoop框架下的情报分析大数据调度超时预测方法]," *Computer Science* [计算机科学], Supp. 1, 2014; Li Chongdong [李崇东], "Study on the Construction of the Military Decision-Making System Based on Large Data Support [基于大数据支持的军事决策系统构建研究]," *Software Engineering* [软件工程], Vol. 3, 2016.

<sup>81</sup> Wang Shoubiao [王寿彪], Li Xinming [李新明], and Liu Dong [刘东], "Concept Association Mechanisms and Model Structures on Equipment System of Systems with Big Data [大数据与装备体系的概念关联机理和模型结构]," *Journal of the Chinese Academy of Electronics and Information Technology* [中国电子科学研究院学报], Vol. 5, 2016.

<sup>82</sup> Fu Zhong-li [傅中力], Zhuang Huang [张煌], and Li Po [李坡], "National Security and Military Strategy Selection in an Era of Big Data [大数据时代的国家安全与军事战略选择]," *National Defense Science and Technology*, Vol. 2, 2013.

позволить командирам PLA иметь более чёткое представление о месте ведения боевых действий и уверенно определять удалённые цели, несмотря на отрывочную информацию. Проекты в области больших данных и искусственного интеллекта были задействованы для анализа входных данных со спутников, визуальных, радиолокационных, инфракрасных, гидролокаторных и других датчиков, а также данных общедоступной разведки.

Самый амбициозный из этих проектов был направлен на интеграцию данных от различных типов датчиков для обнаружения и идентификации целей. В частности, проект стремился использовать спутниковые изображения в режиме онлайн в качестве набора данных и интегрировать по крайней мере пять форм данных: данных о доставке в режиме онлайн (включая текст) с высоким разрешением или мультиспектральные камеры, инфракрасного порта, радара с синтезированной апертурой и информацию об электромагнитном излучении для автоматического отслеживания кораблей. Система должна определять местоположение судна с помощью 85- процентной точности и отслеживать пути навигации корабля с 80- процентной точности<sup>83</sup>. Другой план предусматривал создание прототипа программы, которая могла бы автоматически интегрировать фрагментарные данные с береговых, морских, воздушных и космических радаров; фотоэлектрических датчиков; других датчиков и человеческих разведывательных источников для определения целей в средних и дальних морях<sup>84</sup>. Исторический анализ больших данных также важен при идентификации целей для PLA. Например, радарная бригада Западного театра военных действий PLA использовала систему аналитических баз данных для сравнения характеристик неопознанной воздушной активности с аналогичными случаями в прошлом, что позволило быстро идентифицировать неопознанные самолёты<sup>85</sup>.

---

<sup>83</sup> "Военно-морские исследования: технология добычи спутниковых данных на основе больших данных," Информационная сеть закупок всего военного оружия и оборудования, 1 августа 2016 года. Аналогичное, но, по-видимому, не обеспеченное финансированием предложение от 2016 г. см. в разделе "Использование больших данных и машинного обучения для анализа и прогнозирования из нескольких информационных Sources-315020301", Вся информация о приобретении военного оружия и техники, 28 августа 2016 г. Эта система должна была быть способна отслеживать изменения на поле боя и намерения противника и предсказывать изменения и перемещения цели. Он собирался использовать по меньшей мере шесть методов для идентификации шаблонов целей, четыре метода сбора целевой информации, четыре алгоритма для интеграции целевой информации, три метода для помощи в нацеливании и три метода визуализации поля [海军预研-基于大数据的卫星信息数据挖掘技术], *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], August 1, 2016. For a similar but apparently unfunded proposal from 2016, see "Using Big Data and Machine Learning to Analyze and Forecast from Multiple Information Sources-315020301 [信息系统-315020301-大数据背景下基于深度学习的多源情报分析与预测技术], *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网].

<sup>84</sup> "Naval Innovation—30201050404—Technology to Provide Information on Targets in the Middle and Far Seas with Big Data [海军创新-30201050404-基于大数据的中远海海上目标信息处理技术], *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], April 25, 2017. For another, possibly related project, see "Naval Advanced Research—Ship Big Data Mining and Processing Technology [海军预研-船舶大数据挖掘与处理技术], *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], August 14, 2016.

<sup>85</sup> Jia Chong [贾冲], "Databases, Assisting High-Quality and Rapid Support [数据库, 助保障优质快速], *Air Force Daily*, May 27, 2016.

Есть много других примеров того, как PLA ведёт проекты C4ISR при поддержке больших данных или с большими данными. Согласно одному китайскому уведомлению, PLA планирует создать систему, использующую машинное обучение на большой библиотеке инфракрасных изображений, визуальных изображений и видеоизображений с частотой 25 кадров в секунду военных кораблей и гражданских судов, чтобы создать метод, позволяющий различать их между собой с 95-процентной точности<sup>86</sup>. Другой проект PLAN направлен на использование изображений со всех удалённых датчиков для создания базы данных изображений, на которой можно подготавливать алгоритмы глубокого обучения для дифференциации различных типов военных кораблей, а затем использовать эти алгоритмы для создания мультисенсорного устройства платформы обнаружения военных кораблей с несколькими датчиками<sup>87</sup>. В идеальных условиях камера с разрешением 2,5 метра должна быть способна определять военные корабли среднего размера с точностью до 95% и только 10% ложных срабатываний. Если разрешение составляет от 10 до 16 метров, платформа должна уметь идентифицировать большие военные корабли (авианосцы) с точностью до 90% и ложными срабатываниями до 15%. Согласно деталям синопсиса одного из таких связанных проектов, PLA стремилась создать установленные на самолётах мультисенсорные и мультиспектральные системы для различения почвы, горных пород, воды и синтетических объектов с точностью до 90%<sup>88</sup>. В синопсисе другого проекта мы обнаружили, что PLA начала небольшое патентное исследование состояния компьютерного зрения, сверхчеловеческого визуального восприятия и отслеживания, онлайн-анализа изображений и видео, а также трёхмерного воссоздания из изображений<sup>89</sup>.

PLA также использует возможности анализа больших данных, чтобы различать военные сигналы. В одном примере PLA попытался использовать частично контролируемое обучение для определения характеристик различных типов электромагнитных сигналов, таких как сигналы от оборудования связи, радаров и навигационных систем. Эта система могла бы идентифицировать характеристики сигнала (и, предположительно, таким образом идентифицировать его) в течение 0,1 доли секунды и моделировать сигналы, которые она получает в

<sup>86</sup> "Naval Innovation—30201050111—Image Processing and Ship Target Recognition Based on Artificial Intelligence [海军创新-30201050111-基于人工智能的图像处理 and 舰船目标识别技术]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], March 31, 2017.

<sup>87</sup> "Naval Innovation—30201050110—Marine Remote Sensing Target Information Mining Technology Based on Deep Learning [海军创新-30201050110-基于深度学习的海洋遥感目标信息挖掘技术]," *Whole Military Weapons and Equipment Purchase Information Net*, March 31, 2017.

<sup>88</sup> "Key Laboratory Fund—6142A010103—New Model of Spectral Matching for Ground Objects Supported by Multisource Remote Sensing Big Data [重点实验室基金-6142A010103-多源遥感大数据支持下的地物光谱匹配新模型研究]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], May 19, 2017.

<sup>89</sup> "Artificial Intelligence Series 3—Patent Analysis of Key Technologies in Computer Vision [人工智能系列3—计算机视觉关键技术专利分析]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], December 23, 2017.

течение 10 секунд<sup>90</sup>. В области идентификации цели PLA пытается разработать технологию, которая обеспечивает высококачественную дифференциацию радиолокационных изображений с синтезированной апертурой с использованием машинного обучения для определения характеристик целей<sup>91</sup>. В отдельном проекте планировалось использовать глубокое обучение для создания метода, который использует несколько фотоэлектрических источников для определения характеристик цели (предположительно для идентификации и дифференциации цели) и движения<sup>92</sup>. PLA также стремилась использовать глубокое обучение для улучшения различения целей с помощью радаров, особенно в связи с изменением формы цели или наличием несовершенных исторических данных о цели<sup>93</sup>. Аналогичным, в подводной области PLAN заинтересован в разработке технологий, которые позволяют использовать глубокое обучение и генерацию моделей для идентификации подводных целей с использованием ограниченных акустических данных<sup>94</sup>.

Наконец, PLA интересуют возможности визуализации данных. Мы обнаружили ссылки в статьях, в которых подчёркивается важность визуализации и интеграции данных в командные платформы. Несмотря на то, что у нас нет доступа к этим статьям, они были написаны Лабораторией моделирования сложных электронных систем Колледжа оборудования, Командным колледжем PLAAF, Университетом Освободительной армии и инженерным факультетом технологических коммуникаций, а также 28-м научно-исследовательским институтом корпорации China Electronics Technology Group. (под непосредственным контролем МПТ).

---

<sup>90</sup> “Fund—61401370502—Research on Signal Feature Extraction Based on Deep Learning [基金-61401370502-基于深度学习的信号特征提取技术研究],” *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], August 1, 2016.

<sup>91</sup> “Key Laboratory Fund—61425030202—Microwave Imaging Automatic Recognition Technology [重点实验室基金-61425030202-微波成像自动识别技术],” *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], May 19, 2017.

<sup>92</sup> “Key Laboratory Fund—61421070104—Optical Information Fusion Technology Based on Big Data [重点实验室基金-61421070104-基于大数据的光电信息融合技术],” *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], May 19, 2017.

<sup>93</sup> “Fund—61404130305—Radar Target Recognition Technology Based on Deep Learning [基金-61404130305-基于深度学习的雷达目标识别技术],” *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], April 11, 2017.

<sup>94</sup> “Naval Innovation—30202021401—Underwater Sound Detection and Recognition Based on Big Data [海军创新-30202021401-基于大数据的水声探测与识别技术],” *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], March 31, 2017. Also see “Fund—61404160301—Underwater Learning Target Recognition Method Based on Deep Learning and Generating Modeling (Key Points) [基金—61404160301—基于深度学习与生成式建模的水中目标识别方法（重点）],” *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], August 1, 2016.

## Оборудование и обслуживание

PLA считает, что большие данные национальной обороны будут полезны как при закупке, так и при обслуживании военной техники. По мнению военных учёных, PLA должна укрепить свою военную технику с помощью анализа больших данных<sup>95</sup>. Они предполагают, что, даже если войны будут вестись с помощью оборудования, программное обеспечение будет решающим фактором в том, будут ли эти системы в конечном итоге эффективными. Что касается закупки оборудования, PLA поручила создать систему моделирования и управления цепочкой поставок, оптимизированную с использованием обратной связи и анализа больших данных, а также программного обеспечения, которое использует большие данные для анализа и визуализации бизнес-процессов<sup>96</sup>. Отдельно китайский военный учёный заявил, что PLAN должен построить интеллектуальную производственную систему для военных кораблей и их компонентов, включая анализ изображений для выявления отказов корпуса и трёх – размерное моделирование, достаточно точное, чтобы служить основой для сборки<sup>97</sup>. Программное обеспечение для моделирования должно добавлять необходимые системы автономно. Эти функции должны повысить как эффективность производства, так и согласованность оборудования на 30%. В более широком смысле, согласно другой статье, PLA предполагала организацию оборонной промышленности Китая с использованием больших данных<sup>98</sup>. Эта инициатива позволит определить потребности бизнес-департаментов в больших данных и создаст модельную систему, которая использует законные и технически осуществимые методы для получения, хранения, анализа и визуализации этих данных. Эта система могла бы помочь с оборудованием, логистикой, мобилизацией, набором персонала, обучением, техническими инновациями и совместной работой гражданского и военного бизнеса.

PLA также возлагает большие надежды на поддержку больших данных для обслуживания оборудования. Например, один проект предусматривает сбор данных по 300 параметрам, включая системную информацию, сбои и записи о техническом

<sup>95</sup> Liu Junjie [刘俊杰], Zhang Wenjun [张文军], and Chen Zhang [陈张], "Impact of Big Data on Military Equipment Construction and Development [大数据对军事装备建设发展的影响]," *Journal of the Military Transportation University* [军事交通学院学报], Vol. 10, 2015.

<sup>96</sup> "Production Process Simulation and Control System Based on Big Data [基于大数据的生产过程仿真和控制系统]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], November 22, 2016; "Autonomous Business Data Analysis and Visualization Platform Software Purchasing Requirements [商业智能数据可视分析平台软件采购需求]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], April 10, 2019.

<sup>97</sup> "Naval Innovation—30205020708—Intelligent Control Technology for Construction Process Based on Big Data Analysis [海军创新-30205020708-基于大数据分析的建造过程智能管控技术]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], March 31, 2017.

<sup>98</sup> "Business-Oriented Big Data Mining Analysis [业务导向的大数据挖掘分析]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], April 21, 2017.

обслуживании, для обеспечения очень раннего предупреждения об отказе оборудования с точностью до 70%<sup>99</sup>. Другой проект был направлен на использование больших данных, искусственного интеллекта и распознавание изображений для выявления трещин в гребном винте двигателя и лопастях рабочего колеса с точностью до 90%<sup>100</sup>.

## Военное здравоохранение и мобилизация

В области военного здравоохранения Главный госпиталь PLA и Пекинский университет совместно учредили первую в Китае Национальную лабораторию прикладных медицинских технологий больших данных. Это произошло под руководством Департамента высокотехнологичной промышленности NDRC и Управления здравоохранения Департамента материально-технического обеспечения Центральной военной комиссии<sup>101</sup>. Отдельно отдел научных исследований специального военного училища планировал использовать большие данные для оценки вероятности травм на военно-спортивных тренировках и улучшить управление рисками<sup>102</sup>.

По крайней мере, два проекта касаются того, как большие данные могут улучшить мобилизационные усилия Китая. Fujian Electrical Engineering Group (Фуцзяньская электротехническая группа) изучала, как Интернет и автоматизированный сбор данных могут улучшить понимание Китаем его мобилизационного потенциала<sup>103</sup>. В другой статье рассматривается Национальное управление гражданской противовоздушной обороны и его планы по завершению строительства инфраструктуры национальной платформы больших данных для гражданской противовоздушной обороны к 2019 году<sup>104</sup>. Это повлечёт за собой разработку стандартов данных и создание центров обработки данных по

<sup>99</sup> “Common—41402050301—System-Level Product Fault Symptom Discovery and Fault Prediction Research Based on Big Data [共用-41402050301-基于大数据的系统级产品故障征兆发现与故障预测研究],” *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], May 23, 2017.

<sup>100</sup> “303060302—Air Force Equipment Pre-Research Innovation—Research on Engine Blade Crack Detection Based on Automatic Image Recognition [303060302-空军装备预研创新-基于自动图像识别的发动机叶片裂纹检测研究],” *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], September 1, 2017.

<sup>101</sup> Wang Junping [王君平] and Luo Guojin [罗国金], “China’s First Medical Big Data Applied Technology Laboratory Established at the PLA General Hospital [国内首家医疗大数据应用技术国家工程实验室在解放军总医院成立],” *People’s Daily*, November 26, 2017.

<sup>102</sup> Wang Wang [王王], “Based on the Risk of Military Sports Training Under Big Data Management Research [基于大数据下的军事体育训练风险管理初探],” *Journal of Military Physical Education and Sports* [军事体育学报], Vol. 1, 2017.

<sup>103</sup> Zhong Jun [钟军], “Exploring the Use of Big Data to Innovate the Investigation of the Country’s Mobilization Potential [关于利用大数据创新国防动员潜力调查工作模式的探索],” *Information Technology and Informationization* [信息技术与信息化], Vol. 9, 2016.

<sup>104</sup> An Chunhong [安春红], “Civil Air Defense Big Data Platform Will Be Completed Before 2019 [人民防空大数据平台2019年前初步建成],” *China National Defense News* [中国国防报], January 10, 2017.

гражданской ПВО на национальном, провинциальном и муниципальном уровнях. По состоянию на июль 2020 года китайские провинции, похоже, продолжают совершенствовать свои системы больших данных гражданской ПВО<sup>105</sup>.

## Обучение и набор

PLA ясно видит последствия анализа больших данных для достижения целей обучения и найма. Согласно статье 96754 отдела военной политической работы, большие данные могут использоваться в политической работе и образовании для отслеживания действий и мыслей солдат, индивидуализации обучения и лучшей оценки образования<sup>106</sup>. В другом случае учебная база в Вооружённых силах Гуанчжоу использовала базу данных, содержащую различные сведения по таким вопросам, как семейное происхождение, хобби и интересы, сообщения в социальных сетях, а также повседневные мысли и поведение для оптимизации курсовой работы по политическим вопросам<sup>107</sup>. Примечательно, что Центральная полицейская академия опубликовала статью, в которой указывается на аналогичное применение больших данных в образовании для мониторинга повседневной деятельности, мыслей и политического воспитания, а также психического и физического здоровья учащихся<sup>108</sup>.

В будущем фактические обучающие мероприятия также могут в некоторой степени управляться и поддерживаться с помощью анализа больших данных. Например, во время проведения учений по противовоздушной и противоракетной обороне отряд флота Южных морей использовал каналы передачи данных корабль – воздух, корабль – корабль и корабль – берег для создания единого «поля битвы данных». Отряд использовал данные, собранные ранее группами по сбору данных, и данные о кораблях, оборудовании и личном составе для оцифровки обстановки на поле боя и обеспечения точной основы для процессов рассмотрения последующих действий<sup>109</sup>. Кроме того, подразделения PLAAF командования Центрального театра

<sup>105</sup> Гражданское Главное управление Противовоздушной обороны провинции Цзянсу, Команда и Информационное бюро “Провинциальный Гражданский Офис Противовоздушной обороны Созывает Конференцию по ‘Исследованию Гражданских Информационных систем Командования ПВО На основе Больших данных”, [省人防办召开《基于大数据支撑的人防指挥信息系统研究》开题评审会], Jiangsu Provincial Civil Defense Bureau, webpage, July 24, 2020.

<sup>106</sup> Ding Jiayou [丁佳友], “Using Big Data in Ideological and Political Education [大数据在思想政治教育中的运用],” *Political Work Studies* [政工学刊], Vol. 8, 2017.

<sup>107</sup> He Weili [何伟理], “Building ‘Big Data’ for ‘Living Thoughts’ [给 “或思想” 建立大数据],” *Warrior News* [战士报], August 31, 2015.

<sup>108</sup> Chai Ruobing [柴若冰] and Ma Guofu [马国富], “Using the Central Police Academy as an Example to Explore the Use of Big Data to Manage a Police Academy [大数据在警察院校警务化管理中的应用探析以中央司法警官学院为例],” *Statistics and Management* [统计与管理], Vol. 5, 2016.

<sup>109</sup> Li Youtao [黎友陶], “‘Data Wars’ Play Out in Distant Ocean [‘数据战场’ 亮剑远海大洋],” *People’s Liberation Army Daily*, May 19, 2016.

военных действий организовали 129 боевых вылетов с авиабаз на расстояние до нескольких тысяч километров от командных модулей во время парада PLA в августе 2017 года на учебной базе Чжурихэ. Командные модули выполняли обработку больших объёмов данных для определения наиболее оптимальных маршрутов полёта для обеспечения точных и детальных командных операций<sup>110</sup>.

PLA также надеется повысить профессионализм офицеров с помощью больших данных. В одном примере Шицзячжуанская механизированная пехотная школа и Хэбэйский медиауниверситет стремились использовать большие данные для улучшения дистанционного образования, методов обучения и распределения образовательных ресурсов<sup>111</sup>. В статье Колледжа раннего предупреждения PLAAF оценивается, как вооружённые силы США используют большие данные в профессиональной подготовке и то, как PLA может последовать их примеру<sup>112</sup>. Шанхайский кампус Нанкинского политического колледжа также выпустил доклад о больших данных при принятии решений в области военных кадров<sup>113</sup>.

Что касается набора в PLA, то военное ведомство, базирующееся в Хайлине, Хэйлунцзян, например, использовало платформу больших данных для выявления высокообразованных молодых людей подходящего возраста. Это улучшило способность PLA рассылать материалы о наборе персонала целенаправленно посредством социальных сетей через заранее определённые промежутки времени<sup>114</sup>.

## Моделирование и симуляция

Моделирование и симуляция составляют ещё одну область интересов PLA в области больших данных национальной обороны. Например, согласно амбициозной статье Школы информационных систем и управления Национального технологического университета обороны, PLA должна создать «киберфизическую социальную систему» для военного дела<sup>115</sup>. В такой системе создавался бы

<sup>110</sup> “‘Big Data’ Transmits the Air Force’s Strategic Transition [‘大数据’ 透射空军战略转型],” *Science and Technology Daily*, July 31, 2017.

<sup>111</sup> Wang Pingli [王萍丽], Wu Zheng [吴政], and Wu Yinghao [吴英昊], “Researching Problems of Modern Military Distance Education in an Era of Big Data [大数据时代军队现代远程教育发展问题研究],” *Continuing Education [继续教育]*, Vol. 12, 2014.

<sup>112</sup> Tan Shaoying [谈少盈], Li Xiaoping [李小平], and Wang Pinpin [王晶晶], “Practice of the U.S. Military’s Using Big Data for Promoting Military Vocational Education and Their Enlightenments [美军运用大数据促进军事职业教育的做法及启示],” *Journal of the Air Force Early Warning Academy [空军预警学院学报]*, Vol. 4, 2016.

<sup>113</sup> Xing Jinrong [邢金融] and Zhang Yiming [张一明], “Using Big Data to Improve the Effectiveness of Military HR [运用大数据提升军事人力资源管理效能],” *Theoretical Studies on PLA Political Work [军队政工理论研究]*, Vol. 5, 2016.

<sup>114</sup> Liu Jianwei [刘建伟], “In the Era of Big Data, How Can the Work of Armed Forces Departments Be Precisely Upgraded? [大数据时代，武装工作如何精准发力],” *People’s Liberation Army Daily*, April 27, 2017.

<sup>115</sup> Wang Feiyue [王飞跃], “The Coming Revolution in National Defense Weaponry and Systems: From 3-D Printing to Parallel Military Systems [国防装备与系统的未来变革：从3D打印到平行军事体系],” *National Defense Science and Technology [国防科技]*, Vol. 3, 2013.

виртуальный вариант боевого пространства и постоянно совершенствовался бы для обеспечения точности. Это поможет обучить командиров и проверить стратегии или тактику. Он также может быть привязан к подразделениям на реальном поле боя, чтобы отслеживать их передвижения и лучше контролировать их. Такая система могла бы позволить командиру запускать ракеты одним нажатием кнопки и была бы столь же важна в военном деле, как интеллектуальные процессы управления на производстве. Более того, на 24-м заседании Китайской (военной) ассоциации специалистов системотехники его участники обсудили большие данные в контексте боевой оценки и боевых экспериментов, предложив возможную связь между этими концепциями<sup>116</sup>. Наконец, мы обнаружили, что PLA предложила проект в 2017 году на методы исследования и общее состояние области использования ИИ и глубокого обучения в моделировании и симуляции оборудования<sup>117</sup>.

## Кибер-безопасность

Кибербезопасность, по-видимому, является ещё одной важной областью, в которой PLA уделяет большое внимание большим данным. Например, мы обнаружили проект PLA AF, направленный на сбор данных о поведении и данных о безопасности со всех оконечных терминалов в режиме реального времени для выявления угроз<sup>118</sup>. Эта система должна иметь возможность моделировать угрозы, используя онлайн-информацию, а затем идти вверх по течению, чтобы найти источник проблемы. В другом случае PLA стремилась разработать технологии киберзащиты и мультидоменной безопасности<sup>119</sup>. Исследовательский проект PLA 2019 года направлен на использование состязательных процессов машинного обучения для построения алгоритмов, способных идентифицировать вредоносные данные, вводимые в автоматизированные системы, а также другие алгоритмы, которые могут привести к сбоям в работе PLA в эксплуатационных условиях<sup>120</sup>. Для

<sup>116</sup> Liu Xujiang [刘书江], Zhao Cunru [赵存如], and Li Ning [李宁], "Summary of the 24<sup>th</sup> Annual Meeting of the China Systems Engineering Academy Military Systems Engineering Professional Group [中国系统工程学会军事系统工程专业委员会第二十四届学术年会成果综述]," *Military Operations Systems Engineering* [军事运筹与系统工程], Vol. 4, 2014.

<sup>117</sup> "Key Laboratory Fund—61420080104—Application of Artificial Intelligence Based on Deep Learning in Simulation of Army Equipment System [重点实验室基金-61420080104-基于深度学习的人工智能技术在陆军装备体系仿真中的应用研究]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], May 19, 2017.

<sup>118</sup> "303060501—Air Force Equipment and Research Innovation—Air Force Information Network End-Point Security Response and Monitoring Big Data System [303060501-空军装备预研创新-空军信息网络终端安全大数据分析响应与监控系统]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], August 10, 2017.

<sup>119</sup> "315075703 Automated System Data and Algorithm Safety Evaluation Technology [315075703智能系统数据与算法安全检测技术]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], February 14, 2019.

<sup>120</sup> "Key Laboratory Fund—61421030206—Research on Big Data Security and Privacy Protection Technology [重点实验室基金-61421030206-大数据安全及隐私保护技术研究]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], May 19, 2017.

повышения безопасности оборудования PLA заказала создание программного обеспечения, которое сможет определять, когда были подделаны схемы или микросхемы, анализируя их изображения<sup>121</sup>.

В более широком смысле PLA чётко отдаёт приоритет кибервойне и интеграции данных, полученных в результате этих усилий, с другими данными из связанных или различных областей боевых действий. Наиболее ярким примером этого является создание PLA в декабре 2015 года Сил стратегической поддержки НОАК (PLASSF). PLASSF отвечает за интеграцию кибер-данных с ЕМ и информацией о космической войне<sup>122</sup>. Это уникальная китайская договорённость с неоднозначными шансами на успех. Тем не менее, создание такой организации подчёркивает, насколько серьёзно и насколько смело Пекин думает об интеграции больших данных в стране в поддержку совместных военных операций.

### **Next Up: переход от информатизации к интеллектуализации**

В ближайшие годы Китай надеется овладеть большими данными национальной обороны в достаточной степени, чтобы контролировать или даже доминировать в информационной среде ведения войны против великих держав. Однако это только начало. Большие данные национальной обороны также лежат в основе долгосрочных усилий Пекина, сформулированных президентом Си в его речи на 19-м съезде партии, по превращению в конечном итоге к 2030 году в глобальный центр искусственного интеллекта. Это включает создание интеллектуальной армии<sup>123</sup>.

По правде говоря, Пекину ещё предстоит пройти долгий путь, чтобы реализовать огромный потенциал, который он видит в автоматизированных средствах поддержки принятия решений. Тем не менее, во многих областях Пекин активно рассматривает возможность применения технологии искусственного интеллекта для военных приложений<sup>124</sup>. PLA хочет создать цифрового штабного

<sup>121</sup> “31512050403 Machine Learning Technology to Detect Trojan Horses in Integrated Circuits [31512050403基于机器学习的集成电路硬件木马版图检测技术],” *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], April 30, 2019.

<sup>122</sup> For more on the PLASSF, see Kevin L. Pollpeter, Michael S. Chase, and Eric Heginbotham, *The Creation of the PLA Strategic Support Force and Its Implications for Chinese Military Space Operations*, Santa Monica, Calif.: RAND Corporation, RR-2058-AF, 2017.

<sup>123</sup> См., например, редакционный отдел китайской военной науки, “Резюме семинара по игре между AlphaGo и Lee Sedol и интеллигенции военного командования и принятия решений”, [围棋人机大战与军事指挥决策智能化研讨会观点综述], *China Military Science* [中国军事科学], April 2, 2016.

<sup>124</sup> Авторитетный список и анализ зарождающихся военных возможностей Китая в области ИИ см. в документе Эльза Б. Каня, Сингулярность поля боя: искусственный интеллект, военная революция и будущая военная мощь Китая, Вашингтон, округ Колумбия: Центр новой американской безопасности, 28 ноября 2017 года. Для более широкого изучения планов ИИ Китая во всех областях, включая военные, см. Jeffrey Ding, Расшифровка китайской мечты ИИ: контекст, компоненты и последствия стратегии Китая по руководству миром в ИИ, Оксфорд, Великобритания: Институт будущего человечества, Оксфордский университет, март 2018 года.

офицера, который будет интегрирован в её Систему управления совместными операциями для оперативной интеграции и информационного планирования<sup>125</sup>. Однажды этот цифровой штабной офицер сможет давать рекомендации командованию по повышению оперативности и точности принятия решений в соответствии с программой Deep Green Агентства перспективных исследовательских проектов Министерства обороны США. Недавние исследовательские проекты PLA начали предпринимать шаги по созданию алгоритмов, которые могут помочь командирам использовать данные для принятия более эффективных решений на поле боя. Два таких проекта нацелены на использование алгоритмов машинного обучения для извлечения оптимальных тактик или стратегий из данных, созданных в ходе прошлых операций<sup>126</sup>. Другой проект стремился достичь аналогичного результата с помощью алгоритмов дуэлей, которые оттачивали бы их навыки, играя в итеративные реалистичные тактические симуляторы против друг друга<sup>127</sup>. PLA также изучает возможности улучшения военных игр с помощью интеллектуальных противников, использующих ИИ, и реалистичного обучения<sup>128</sup>. Чтобы реализовать эти возможности, PLA стремится разработать меньшее, более лёгкое и быстрое оборудование, чтобы его компьютеры могли проводить сложные вычисления в оперативных условиях достаточно быстро, чтобы китайские командиры могли наблюдать, ориентироваться, принимать решения и действовать (OODA) быстрее, чем их оппоненты<sup>129</sup>. Компания продолжает изучать планы по совершенствованию обработки естественного языка, автоматического перевода и анализа платформ социальных сетей<sup>130</sup>, также исследует когнитивное радио, использование ИИ для защиты от глушения и лучшего подавления частот противника<sup>131</sup>.

Пекин считает, что ИИ имеет важное значение и для военной техники. Он изучает более умные крылатые ракеты с ИИ; высокоавтономные беспилотные

<sup>125</sup> Kania, 2017.

<sup>126</sup> "Space Flight System Modeling Major Laboratory Fund—6142002302—Autonomous Systems to Identify Optimal Paths to Victory Based on Shifting Situations [航天系统仿真重点实验室基金—6142002302—基于事件演化的智能化制胜规则寻优方法]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], November 25, 2018; "3151120201 Autonomous Technology to Generate, Optimize, and Evaluate Military Knowledge Based on Machine Learning [3151120201基于机器学习的军事知识生成、演化与评估技术]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], April 30, 2019.

<sup>127</sup> "31505550302 Learning from and Modeling Autonomous Adversarial Games [31505550302面向智能博弈的对抗策略建模与学习技术]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], February 14, 2019.

<sup>128</sup> Hu Xiaofeng, He Xiaoyuan, and Tao Jiuyang, "Alpha Go's Breakthrough and the Deductive Challenge of War Games [AlphaGo的突破与兵棋推演的挑战]," *Science and Technology Review* [科技导报], Vol. 35, No. 17, 2017, p. 58.

<sup>129</sup> "31511020105 Embedded Platforms for Smart Computing [31511020105装备嵌入式智能计算平台技术]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], December 23, 2017. Note that the description of this project explicitly references the OODA loop.

<sup>130</sup> "Artificial Intelligence Project 2—Analysis of Critical Patents in Natural Language Processing [人工智能系列2——自然语言处理关键技术专利分析]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], December 23, 2017.

<sup>131</sup> Kania, 2017.

летательные аппараты; скопление беспилотных летательных аппаратов; высокоавтономные внедорожные вооружённые и невооружённые беспилотные наземные машины; зенитно-артиллерийские или ракетные лазерные комплексы, способные автономно обнаруживать, сопровождать и поражать маловысотные цели, а также высокоавтономные надводные и подводные беспилотные аппараты дальнего действия<sup>132</sup>. Наконец, Пекин рассматривает возможность использования операционных систем и микросхем для систем искусственного интеллекта, обеспечения безопасности автономных систем, командного управления пилотируемыми и беспилотными системами, а также автономно скоординированных операций со многими типами дронов<sup>133</sup>. Управление большими данными национальной обороны и их использование являются первыми в серии сложных шагов на пути к развитию всех этих и других систем искусственного интеллекта.

## ГЛАВА ПЯТАЯ

### Заключительные мысли и будущее

Проводя исследования, Пекин явно возлагает большие надежды на аналитику больших данных. Китайские лидеры разделяют общее видение того, что использование больших данных будет играть важную роль, помогая Китаю стать настоящей великой державой. Однако они также считают, что стать мировым лидером в информационную эпоху может только Китай. Пекин стремится использовать технологии и навыки анализа больших данных, чтобы к 2030 году стать глобальным центром искусственного интеллекта или, как это называют китайские лидеры, интеллектуального принятия решений.

Между тем, судя по двум нашим тематическим исследованиям, мы считаем, что Пекин глубоко понимает ценность интеграции больших данных в правительственную и военную деятельность. Будь то в области предупреждения

<sup>132</sup> Kania, 2017. Also see Abhijit Singh, "Is China Really Building Missiles with Artificial Intelligence?" *The Diplomat*, September 21, 2016; "Army Advanced Research—0243—Independent Coordination for Multiple Drones [陆军预研-0243-无人机多机自主协同技术]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], July 27, 2016; Jeffrey Lin and P. W. Singer, "China's New Military Robots Pack More Robots Inside (Starcraft-Style)," *Popular Science*, November 11, 2014; Jeffrey Lin and P. W. Singer, "New Chinese Laser Weapon Stars on TV," *Popular Science*, November 25, 2015; "Major Laboratory Fund—61422150101—AI Navigation Methods for Unmanned Maritime Navigation [重点实验室基金-61422150101-面向水中无人航行器的人工智能方法]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], May 19, 2017; "An Age in Which We Reach a Depth of 1,000 Meters," *IQIVI*, November 19, 2016.

<sup>133</sup> State Council, "Plan for Developing a New Generation of AI [新一代人工智能发展规划]," July 8, 2017; "Major Laboratory Fund—61423011001—Distribution and Resource Optimization and Path Planning for Multiple Drone Cooperative Missions [重点实验室基金-61423011001-异构多无人机协同任务分配、资源优化和路径规划系统]," *Whole Military Weapons and Equipment Purchase Information Net* [全军武器装备采购信息网], May 19, 2017.

преступности и контроля за поведением своих граждан или в повышении способности PLA проводить совместные военные операции, существует впечатляющее разнообразие программ в области анализа больших данных, которые были либо предложены, либо реализуются в Китае. Кроме того, как демонстрирует этот отчёт, открытые китайские первоисточники обширны и иногда довольно подробно описывают конкретные планы инициатив в области больших данных. Мы действительно сталкивались с несколькими ситуациями, когда какие-либо детали были удалены из их исходных веб-источников, либо запрашиваемая информация была отклонена по соображениям конфиденциальности. В будущем мы будем стремиться дополнить наши несекретные выводы секретно полученной информацией, чтобы получить более полную картину усилий Китая по работе с большими данными в этих областях.

Основываясь на двух наших тематических исследованиях, мы считаем, что дальнейшие исследования следующих вопросов позволят лучше понять участие Китая и его инвестиции в аналитику больших данных:

Как Китай мог бы извне использовать свои развивающиеся возможности больших данных для мониторинга и контроля своего населения? Что Пекин может узнать о потенциальных противниках для принуждения или иного влияния на внутренние события в этих странах, используя большие данные? В течение многих десятилетий Пекин применял тактику «единого фронта» против тайваньского народа, чтобы разделить его, пытаясь сделать его более приемлемым для требований Китая по вопросу о суверенитете<sup>134</sup>. Степень, в которой новые возможности Китая по анализу больших данных могут ещё больше расширить действия единого фронта, как в мирное, так и в военное время, неясны. Аналогичным образом, в апреле 2015 года китайские спецслужбы успешно взломали Управление персонала США (OPM), предоставив Пекину огромные массивы данных о персонале США, которые также могут оказаться полезными<sup>135</sup>. Учитывая недавние разоблачения разведывательного сообщества о том, что Россия использовала социальные сети для того, чтобы попытаться повлиять на исход президентских выборов в США, возможно, что Пекин аналогичным образом будет использовать эти личные данные для запуска целевых пропагандистских и дезинформационных кампаний во время будущих выборов. В равной степени возможно, что Пекин может использовать взломанные данные OPM для поддержки военных операций в будущем вооружённом конфликте с США.

---

<sup>134</sup> See, for example, Lauren Dickey, "Can Taiwan Defend Identity from China's United Front Tactics?" *News Lens*, blog post, April 4, 2018.

<sup>135</sup> Хотя Китай конкретно не упоминается как исполнитель атаки, официальный отчет об уязвимостях OPM, позволяющих атаке, можно найти в Комитете по надзору и правительственной реформе, Палата представителей США, нарушение данных OPM: как правительство поставило под угрозу нашу национальную безопасность для более чем поколения, Вашингтон, округ Колумбия, отчет большинства сотрудников, Анализ роли Китая во взломе см., например, в статье Брендана И. Кёнера "Внутри кибератаки, которая потрясла правительство США" *Wired*, 23 октября 2016 года.

Как оцениваются развивающиеся возможности Китая в области анализа больших данных по сравнению с Соединёнными Штатами в военной сфере? Детальная чистая оценка PLA по сравнению со способностью вооружённых сил США использовать большие данные национальной обороны позволит оценить существующую инфраструктуру и технологии, а также уровень подготовки и опыта персонала в обеих странах, чтобы использовать большие данные в боевых действиях. В рамках этого анализа также будут рассмотрены и сопоставлены будущие инвестиции в аналитику больших данных для определения будущих тенденций и вероятных сценариев до 2030 года.

Помимо этих возможных последующих исследований для наших двух тематических исследований, существует как минимум три других отдельных области потенциального интереса в области анализа больших данных:

Как Китай использует анализ больших данных для поддержки своих процессов НИОКР? Как известно, принятие решений, стоящих за процессами НИОКР в Пекине, непрозрачно. Лучшее понимание возможностей анализа больших данных, интегрированных в процесс НИОКР в Китае, может пролить свет на факторы, влияющие на инвестиционные решения и определяющие их. Это также могло бы позволить Соединённым Штатам лучше определять типы систем PLA в НИОКР.

Как Китай оценивает технологические или экономические тенденции с помощью аналитики больших данных? Стремление Китая создавать собственные инновационные технологии и привлекать мировой бизнес с помощью своей краеугольной инициативы «Один пояс, один путь» почти наверняка выиграет от анализа больших данных. Оценка того, как анализ больших данных используется при принятии решений в области экономики и НИОКР, могла бы дать Соединённым Штатам большее понимание аргументов, лежащих в основе этой политики, и позволить Вашингтону либо дополнять их, либо конкурировать с ними.

Как китайское руководство относится к государственным услугам с использованием больших данных и их последствиях для внутренней стабильности? Общей темой в общедоступной китайской литературе по анализу больших данных является необходимость использования этой возможности для расширения государственных онлайн-услуг. Эта концепция, известная как услуги электронного правительства, также снижает вероятность социальных волнений, которые могут угрожать режиму и внутренней стабильности.

Стремление лидеров Китая к созданию возможностей искусственного интеллекта (ИИ) в целях выполнения ряда гражданских и военных функций начинается с освоения аналитики больших данных – использования компьютеров для анализа больших наборов данных. Исследование, проведённое авторами,

показывает, что Китай активно работает над тем, чтобы стать мировым лидером в области анализа больших данных в рамках своего плана по достижению статуса великой державы; действительно, президент Си Цзиньпин заявил, что Китай должен стать глобальным центром ИИ к 2030 году.

В своих усилиях Пекин руководствуется национальной стратегией по работе с большими данными, которая включает в себя экономические, военные, полицейские и разведывательные функции. Авторы обнаруживают, что Пекин уже использует аналитику больших данных для опроса населения страны и повышения ее военного потенциала. Усовершенствования в области анализа больших данных поддержали Пекин в мониторинге и контроле над своими гражданами, особенно над этническими меньшинствами.

## **Искусственный интеллект в области юриспруденции**

- Машинное обучение,
- Искусственный интеллект,
- Natural Language Processing

### **Статья 1**

#### **Введение**

Тема искусственного интеллекта сегодня – одна из самых обсуждаемых. Перспектива «обеспечить монополию в сфере искусственного интеллекта и стать властелином мира» заставила всех соревноваться в данной области. IT-гиганты, финансовые компании, бизнес-аналитики, университеты и научное сообщество предлагают собственное видение инструментов и методологии решения основных задач при их создании. Однако результаты удручают, особенно в сфере LegalTech.



Что такое настоящий LegalTech, а что скрывается под громкими рекламными слоганами? Почему никому из IT-разработчиков не удалось создать действительно прорывной и функциональный продукт, близкий к цифровому юристу? Какой подход позволил нам решить данные задачи и существенно приблизиться к созданию настоящего юридического искусственного интеллекта?

Какова роль практикующих юристов в процессе разработки инструментов автоматизации?

В данной статье мы хотим поделиться с Вами результатами многолетних исследований в области искусственного интеллекта и предоставить ответы на эти вопросы.

*Disclaimer: мы не критикуем существующие инструменты, а говорим о том, что для решения обозначенного круга задач требуется качественно новый подход к разработке.*

## 1. LegalTech в России

Наша компания более 15 лет успешно работает на рынке консалтинговых и юридических услуг. Обладая значительным юридическим опытом решения самых нестандартных кейсов и сложных проектов в России и за рубежом, мы посвятили не один год научному исследованию и практическим разработкам в области цифровых технологий и перспектив их применения в профессии.

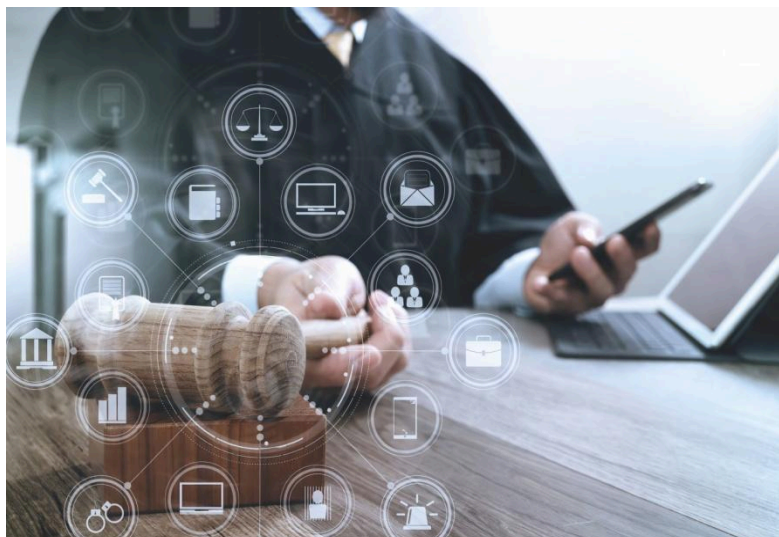
Несколько лет назад мы задались вопросом: почему юридический рынок обделён инструментами автоматизации?

Анализ предлагаемых продуктов позволил прийти к следующим результатам.

Мы, как профессиональные юристы, активно следим за развитием сферы LegalTech и за решениями, которые предлагаются на рынке в качестве прорывных продуктов, способных, по заявлениям создателей, изменить традиционное представление о работе юриста.

Но на самом деле в качестве LegalTech на отечественном рынке распространяются решения, которые очень далеки от содержательной автоматизации юридической функции и позволяют решать локальные задачи, не связанные с творческой и экспертной юриспруденцией.

В то время, как инженеры активно применяют в повседневной работе решения, выполняющие сложные расчёты и рутинные операции, а сотрудники финансового сектора используют цифровые платформы для построения финансовых моделей и оценки рисков, все, что есть у юристов – чуть более продвинутые поисковые сервисы и шаблоны документов.



Весь доступный инструментарий для юриста сегодня это:

- конструкторы документов, работающие на основе типовых и унифицированных шаблонов, в которых любое отклонение от формы требует ручной правки;
- сервисы проверки контрагента, осуществляющие агрегацию общедоступной информации из публичных реестров (ЕГРЮЛ/ЕГРИП, Федресурс, КАД и др.), которые редко позволяют найти ценную информацию;
- системы подбора судебной практики и справочно-правовые системы, осуществляющие базовый поиск, по ключевым словам, фразам, тегам в

открытой базе судебных решений и НПА, которые предоставляют все документы, содержащие искомое слово без учёта контекста и др.;

➤ системы управления проектами, задачами и документами (различные BPM/ERP/ECM-системы, заточенные на автоматизацию биллинга, учёт времени и контроль за ресурсами).

Данные инструменты ни на шаг не приближают нас к автоматизации творческой и экспертной юриспруденции. Они, безусловно, облегчают работу юриста, но только в вопросах поиска информации, а не в её интеллектуальной обработке с точки зрения юридической логики.

Практикующие юристы высокой квалификации согласятся с нами, что если бы можно было предлагать клиентам шаблонные договоры, в которые встроены актуальные даты, суммы и наименования объектов, то профессии юриста уже бы не было. Ценность юриста заключается в его способности предвидеть ситуацию на несколько шагов вперёд и предлагать нестандартные решения в пользу клиента с минимальными рисками и издержками с точки зрения права.

Мы однозначно можем сказать, что при существующем уровне развития технологий юридический рынок в России (и, скорее всего, в мире) не имеет полноценных решений, способных заменить юриста даже начальной квалификации и автоматизировать хоть в сколько-нибудь значимой части юридическую функцию.

## **2. Особенности предметной области**

Прежде чем перейти к анализу технологической стороны вопроса необходимо понять особенности предметной области. Работа любого юриста-эксперта связана с документами. Данные и документы разнородны и имеют собственные отличительные особенности.

Задача профессионального юриста при работе с документами заключается в том числе в правовой оценке их содержания, квалификации отношений между субъектами, выявлении правовых и финансовых рисков для представляемой стороны и выработке предложений по их минимизации или в регламентации существующих отношений между хозяйствующими субъектами.

Например, к юристу обратился клиент, который столкнулся с недобросовестным поведением контрагента и требует защитить его интересы.

Задачами юриста являются:

- запросить и (или) собрать необходимые документы (договоры, акты, письма и все иные материалы, имеющие отношение к делу) и установить значимые юридические факты;
- оценить их под призмой действующего законодательства и практики его применения (понять правовую природу отношений и определить круг правовых норм, подлежащих применению);
- предложить варианты решения проблемы и собственные рекомендации, после – реализовать выбранный вариант.

Это только один из примеров, с которыми сталкивается юрист ежедневно. При последовательном решении каждой отдельно взятой задачи проявляется профессионализм и экспертные навыки, которые формируются по мере работы юриста и накопления опыта.

Одна из значимых компетенций юриста – это умение видеть в письменных документах все юридические факты, выделять наиболее значимые и соотносить их с нормами права для поиска возможных решений.

Именно поэтому одной из ключевых и первостепенных задач, которую необходимо решить для создания действительно функционирующих инструментов автоматизации юридической работы, является обучение машины смысловому пониманию текста на уровне юриста-профессионала.

Речь идёт о полноценном семантическом анализе юридических текстов.

При обращении к вопросам автоматизации юридической функции и создания полноценного юридического искусственного интеллекта мы считаем, что без применения глубоких лингвистических технологий эти задачи решить не получится.

Это, прежде всего, связано с необходимостью научить программные инструменты понимать не только отдельные сущности (категории) в тексте, но и анализировать текст, выделять все возможные смыслы и проводить логические взаимосвязи в его содержании.

В качестве подтверждения данного тезиса приведём следующие аргументы.

В первую очередь, при анализе документа юрист оценивает его содержание с точки зрения смыслов, которые в нем содержатся. Например, в тексте большинства уставов обществ с ограниченной ответственностью имеется следующий пункт, дублирующий п. 3 ст. 21 Федерального закона № 14-ФЗ от 08.02.1998 г.: *«Доля участника Общества может быть отчуждена до полной её оплаты только в той части, в которой она уже оплачена»*.

Применяя экспертные юридические знания, мы можем извлечь следующие смысловые блоки из данного предложения:

- У участника Общества есть доля.
- Доля участника общества может быть отчуждена.
- Доля участника общества может разделяться на части.
- Часть доли может быть отчуждена.
- Участник Общества вправе осуществить отчуждение доли или части доли участника Общества.
- Доля участника Общества оплачивается участником.
- Доля участника Общества может быть оплачена не полностью.
- Часть доли участника Общества может быть оплачена.
- Часть доли участника Общества может быть не оплачена.
- Отчуждение неоплаченной части доли участника Общества запрещено.
- Отчуждение оплаченной части доли участника Общества разрешено и др.

Подобный уровень детализации смыслового содержания документов с помощью машинных инструментов невозможно добиться без воссоздания юридической «картины мира» путём разработки экспертных семантических концептов, созданных в тесном взаимодействии с погруженными в предметную область (как в теорию, так и в практику) специалистами.

Кроме того, с точки зрения внутренней структуры, документы, используемые в юриспруденции, могут быть классифицированы следующим образом:

1. высокоструктурированные документы, имеющие установленную законом строгую форму и упорядоченное содержание (выписки и справки из публичных реестров, документы на бланках строгой отчётности, управленческая документация по ОКУД и др.);

2. слабоструктурированные документы, имеющие, как правило, шаблонную форму, но содержащие некоторые творческие элементы (банковские выписки и др.);

3. неструктурированные документы, не имеющие однородной формы и содержания и характеризующиеся высоким уровнем уникальности содержания (договоры, корпоративные акты, процессуальные документы, юридические заключения, меморандумы и др.).

И если для машинной обработки высокоструктурированных документов сложные лингвистические решения не требуются, то для слабоструктурированных и неструктурированных документов, которых в области права большинство,

технологии NLP (Natural Language Processing) являются единственным инструментом, способным справиться с данной задачей.

Итак, определив приоритетные задачи для создания юридического искусственного интеллекта, мы приступили к анализу рынка и тестированию существующих инструментов NLP. Результаты исследования представлены в Главе 2 настоящей статьи.

## 2.1. Чат-боты и цифровые ассистенты



Пару слов также нужно сказать про ставших недавно популярными цифровых помощников, чат-ботов, ассистентов и т. п. Безусловно, с приходом Alexa, Siri и Алисы множество аспектов бизнеса и нашей повседневной жизни кардинально поменялись:

- подавляющее большинство задач планирования («поставь встречу в календарь») и поиска информации («найди год выпуска фильма») решается без участия человека;
- наверное, самый большой эффект на себе пока ощутила сфера поддержки клиентов, где личное общение со специалистом все чаще становится доступным только для премиального сектора.

В свете такого взрывного роста возникает большой соблазн создать робота-юриста (как end-to-end решение):

- записать ответы юриста на ТОП-100/1000 самых частых вопросов, составить подробный FAQ, описать все возможные жизненные ситуации и т. п.;
- «загрузить» в нейронную сеть всю имеющуюся судебную практику, судебные решения;

- как результат – получить робота, который сможет (например, при помощи deep learning) соотнести запрос пользователя с ответом юриста (судебным прецедентом, решением), который был дан ранее в похожей ситуации.

Такая идея далеко не нова, но все попытки её реализации за всю мировую историю права неизменно заканчивались неудачей.

Причина в следующем: практикующий юрист почти каждый день сталкивается с новой уникальной задачей, которая требует творческого подхода.

В связи с этим технологии чат-ботов и ассистентов мы не расцениваем как элемент LegalTech/Legal AI, поскольку они имеют сугубо опосредованное отношение к автоматизации юридической функции.

Наше субъективное мнение: применение технологий цифровых ассистентов, чат-ботов в сферах экспертной деятельности – крайне рискованное мероприятие:

- ущерб от некорректных действий традиционного чат-бота, например, в сфере поддержки клиентов является номинальным: в случае некорректной работы можно получить недовольного клиента, плохой отзыв и т.п.;
- если ошибается чат-бот в сфере, где необходимы экспертные знания (например, юриспруденция, медицина, строительство) – ущерб может быть непредсказуемым и фатальным.

При этом не составляет особого труда предугадать логику законодательного регулирования. Если цифровые ассистенты и чат-боты будут допущены до экспертной сферы, то, скорее всего, на уровне закона будет установлено, что за все рекомендации и действия ассистентов и чат-ботов их разработчики несут полную ответственность.

### **3. Анализ существующих подходов и инструментов**

Исследованиям вопросов в области лингвистики и технологий NLP мы посвятили большое количество времени в том числе в рамках рабочих встреч и обсуждений подходов с представителями научного сообщества.

Нельзя не оценить их вклад в развитие инструментов обработки текста, которые в настоящее время показывают хорошие практические результаты.

Мы благодарны представителям научного сообщества за бесценный опыт, которым они поделились с нами и внимание, проявленное к нашим разработкам. Прежде всего мы имеем в виду следующие коллективы:

➤ ИТМО (Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики) и, в частности, Руководителя международной лаборатории «Интеллектуальные методы обработки информации и семантические технологии», доцента факультета программной инженерии и компьютерной техники Дмитрия Муромцева, а также его коллег — Любовь Ковригину и Ивана Шилина: под руководством Д. Муромцева был адаптирован для русского языка синтаксический парсер, созданный в Стэнфордском университете для применения на англоязычных текстах (подробнее о тестировании парсера Stanford – в разделе 3.2.3.3.);



**УНИВЕРСИТЕТ ИТМО**

➤ Лабораторию компьютерной лингвистики ИППИ РАН им. А.А.Харкевича (Институт проблем передачи информации Российской академии наук) и, в частности, научных сотрудников Леонида Иомдина и Ивана Рыгаева, под руководством которых был создан синтаксический парсер ЭТАП (и версия ЭТАП-4), применяемый для русского языка (подробнее о тестировании ЭТАП — в разделе 3.2.3.2.).



В России направление NLP развивается во многом благодаря энтузиазму и многолетней работе учёных, осуществляющих разработки в области процессинга русскоязычного текста в условиях ограниченного бюджета, а также отсутствия частных и государственных инвестиций, способных финансировать масштабные исследовательские проекты.

Несмотря на эти обстоятельства, данными научными коллективами были достигнуты огромные практические результаты в области процессинга текста на русском языке, которые заслуживают уважения.

По нашему мнению, обозначенная проблема отсутствия финансовой поддержки и инвестиций в научные разработки технологий NLP в России является одним из факторов, сдерживающих развитие данной области. Государственная поддержка таких проектов позволила бы совершить существенный прорыв в области семантики и синтаксического парсинга, а также достигнуть огромных результатов.

Далее мы предлагаем рассмотреть существующие подходы и инструменты в области обработки текста на русском языке с практической стороны.

### 3.1. Существующие подходы и коммерческие продукты

В области NLP на отечественном рынке присутствует ряд коммерческих решений, которые предлагают универсальный функционал по процессингу русскоязычных текстов. Однако среди данных продуктов отсутствуют профильные решения, ориентированные на юристов и способные удовлетворить потребности предметной области в полном объёме.

Совокупно данные продукты могут быть разделены на две группы. Первая группа представляет собой решения, функционал которых обеспечивается за счёт многочисленного набора правил, которые формулируются лингвистами и специалистами из предметной области. Наиболее известными системами являются решения от АBBYY, Pullenti, Megaputer и ряда других др.

Разработчиками данных решений предлагается проведение лингвистического анализа неструктурированных текстов посредством выделения именованных сущностей (Named Entity Recognition), применения правил морфологии, синтаксиса, семантики и иных процедур обработки (как правило, такие правила описываются в проприетарном закрытом формате).

Стоит оговориться, что в таких системах могут применяться элементы машинного обучения, но они играют второстепенную роль.

Принципиальный недостаток таких решений кроется в подходе – реализация функционала путём создания отдельных правил приводит к необходимости выработать десятки/сотни тысяч правил для отдельно взятой области, что неизбежно приводит к возникновению противоречий. Классический пример:

- для фразы: «студент проходил обучение в МГУ имени М.В.Ломоносова»;
- правила поиска организаций дадут результат – «МГУ имени М.В.Ломоносова»;
- правила поиска имён (ФИО) дадут результат – «М.В.Ломоносов»;

- для разрешения данной (и каждой похожей) ситуации нужно вручную создавать специальное правило-исключение.

Кроме того, одно изменение или ошибка может потребовать пересмотра значительной части правил.

Вторая группа – решения, основанные на применении нейронных сетей, обученных на корпусе текстов. Наиболее яркими представителями являются DeepPavlov, FRED и др. В отличие от первой группы продуктов использование машинного обучения позволяет уйти от необходимости разработки правил анализа текста и их правки при изменениях в предметной области, однако для подготовки обучающего датасета требуется профессиональная разметка сотен тысяч образцов документов каждой используемой категории.

В настоящее время существующие модели предобучены на корпусе текстов из общедоступных источников: художественной литературы, текстов в сети (wikipedia) и др., что не позволяет полноценно использовать их при обработке юридических текстов, обладающих собственной спецификой.

Обучение же на корпусе юридических документов осложнено отсутствием в открытом доступе достаточного количества уникальных образцов в связи с конфиденциальным характером содержания реальных правовых документов.

Именно поэтому в настоящее время нет коммерческих проектов, созданных на основе ML/DL, где в качестве обучающего датасета присутствует достаточное количество юридических текстов.

Тем не менее, многие крупные компании, обладающие собственной обширной базой документов, предпринимают попытки создания инструментов для внутреннего пользования.

### **3.2. Тестирование отдельных инструментов NLP**

Общеизвестен факт, что русский язык во многих аспектах – один из самых сложных языков, особенно когда дело касается профессиональной лексики. Юридические тексты объединяют в себе не только специфическую терминологию, но и формализм, сложную синтаксическую структуру, характеризующуюся наличием множества оборотов (сложносочиненных и сложноподчиненных предложений, причастных и деепричастных оборотов и др.).

Исследуя и подбирая инструменты для решения задачи процессинга юридических текстов на русском языке, мы столкнулись с рядом проблем. Для понимания сложности задачи приведём два примера предложений:

1. *Стоимость товара составляет десять тысяч рублей.*
2. *Согласно пункту 4 Правил безвозмездные целевые взносы предоставляются субъектом оптового рынка на цели выделения из соответствующих бюджетов субъектов Российской Федерации субсидий на возмещение гарантирующим поставщикам, реализующим электрическую энергию (мощность) покупателям на розничных рынках, расположенных в территориально изолированных технологических системах и (или) на территориях, технологически не связанных с Единой энергетической системой России и технологически изолированными территориальными электроэнергетическими системами, а также гарантирующим поставщикам (энергосбытовым (энергоснабжающим) организациям), реализующим электрическую энергию (мощность) покупателям на розничных рынках, расположенных на территориях неценовых зон оптового рынка, недополученных доходов в связи с доведением цен (тарифов) на электрическую энергию (мощность) до базовых уровней цен (тарифов) на электрическую энергию (мощность) в соответствующем периоде регулирования в соответствующем субъекте Российской Федерации. (Решение Верховного Суда РФ от 22 марта 2019 г. № АКПИ18-1182)*

Очевидно, что предложения по типу второго примера чаще используются в юридических текстах, чем по типу первого, что и порождает проблемы в реализации процессинга.

Кроме того, для русского языка неприменимы инструменты, созданные для англоязычных текстов. Причина тому кроется в критических различиях в данных языках. Тогда как английский язык является аналитическим, русский язык обладает главным образом свойствами синтетического языка.

Из этого следует ряд принципиальных отличий между ними. Английский язык имеет фиксированный порядок слов, обеспечивающий структурную связность текста, тогда как в русском языке структура формируется при помощи множества грамматических морфем (приставок, суффиксов, флексий).

Для понимания приведём следующий пример: *«В 2019 году совершена притворная сделка, в соответствии с которой имущество было отчуждено в пользу аффилированного лица мужа члена Совета директоров, фиктивный развод с которым состоялся годом ранее».*

В данном предложении придаточная часть может относиться к нескольким сущностям: «аффилированному лицу», «мужу», «члену Совета». В английском языке этот вопрос разрешился бы за счёт близости главной и зависимой частей.

Английскому языку присуще также обязательное наличие в предложении подлежащего и сказуемого, тогда как русский язык характеризуется возможностью пропуска не только одного из главных членов предложения, но и зачастую слов, смысл которых предполагается, применяя такие фигуры речи, как эллипсис (пропуск слов с возможностью контекстуального восстановления).

Например: «На собрании председательствующий представил аудиторское заключение, ревизор – решение суда», в котором повторяющийся глагол «представил» опускается во второй части и реализуется с помощью пунктуации.

Все это свидетельствует о сложности и уникальности русского языка, особенно с учётом особенностей профессиональной терминологии юристов.

Текущий уровень развития NLP, к сожалению, не позволяет нам сформировать лингвистические универсалии, позволяющие совершать сложные логические операции (анализ, синтез, генерация и вывод) над текстами всех областей человеческих знаний.

Более того в настоящее время нет готовых программных решений по процессингу русского языка даже для крайне узких предметных областей по отдельности, не говоря даже о юридических текстах, которые сочетают в себе как профессиональную лексику, так и общеупотребительные конструкции и терминологию из других сфер.

Технологии NLP строятся на трёх основных направлениях лингвистики: морфологии, синтаксисе и семантике.

Поиск решений в области морфологического анализа для русского языка уже не вызывает острых вопросов: есть несколько готовых качественных инструментов в области морфологии, а также открытых библиотек (например, *natasha/yargy*).

С синтаксисом дела обстоят несколько иначе. Мы исследовали множество парсеров, созданных для русского языка и адаптированных к нему. У всех есть свои преимущества и недостатки. Решение задачи семантического анализа юридических текстов также осложнено рядом проблем. Подробнее об этом расскажем ниже.

### **3.2.1. Семантический анализ**

#### **3.2.1.1. DeepPavlov**

Исследования и разработки с применением семантики направлены на решение многих задач NLP: от глубинного машинного перевода до диалоговых систем и программных решений по генерации готовых текстов.

Изучая научные наработки и готовые open-source модели, мы пришли к выводу, что интенсивнее всего семантическая область NLP на русском языке развивается в сфере систем с естественно-языковым интерфейсом, способным принимать вопросы и отвечать на них на естественном языке (Question-answering system, QA-системы). Однако и в сфере QA-систем отсутствуют профильные решения для юристов.

Одна из наиболее эффективных существующих моделей – DeepPavlov. Данная модель основана на Google BERT (и ряде других моделей) и является открытой программной библиотекой разговорного AI для создания виртуальных диалоговых ассистентов и универсального анализа текста.

Для первого теста мы выбрали предложение осложнённое сочинительной, подчинительной связями, перечислениями, уточнениями, аббревиатурами, сокращениями, числами, производными союзами и предлогами: *«В соответствии со ст. 46 Конституции РФ и гл. 24 АПК РФ граждане и организации вправе обратиться в суд за защитой своих прав и свобод с заявлением об оспаривании решений, действий органов государственной власти, органов местного самоуправления, должностных лиц, государственных или муниципальных служащих, в результате которых, по мнению указанных лиц, были нарушены их права и свободы или созданы препятствия к осуществлению ими прав и свобод либо на них незаконно возложена какая-либо обязанность или они незаконно привлечены к ответственности».*

Тест проводился путём поочерёдных вопросов системе:

1. «Чьи действия можно оспорить?»;
2. «Чьи решения можно оспорить?»;
3. «Какие действия можно оспорить?»;
4. «Какие решения можно оспорить?».

О: органов государственной власти, органов местного самоуправления, должностных лиц, государственных или муниципальных служащих

В: Чьи действия можно оспорить?

В соответствии со ст. 46 Конституции РФ и гл. 24 АПК РФ граждане и организации вправе обратиться в суд за защитой своих прав и свобод с заявлением об оспаривании решений, действий

**органов государственной власти, органов местного самоуправления, должностных лиц, государственных или муниципальных служащих** А, в результате которых, по мнению указанных лиц, были нарушены их права и свободы или созданы препятствия к осуществлению ими прав и свобод либо на них незаконно возложена какая-либо обязанность или они незаконно привлечены к ответственности.

Наглядно видно, что вопрос «*Чьи действия/решения можно оспорить?*» не вызывает трудностей у системы: придаточная часть «органов государственной власти...» относится к однородным членам предложения одинаково, вне зависимости от последовательности этих слов в предложении. Очевидно, вопрос предполагает синтаксическую связь «действий», «решений» со словом/словами в форме родительного (притяжательного) падежа и, вероятно, множественного числа.

О: органов государственной власти, органов местного самоуправления, должностных лиц, государственных или муниципальных служащих

В: Чьи действия можно оспорить?

В соответствии со ст. 46 Конституции РФ и гл. 24 АПК РФ граждане и организации вправе обратиться в суд за защитой своих прав и свобод с заявлением об оспаривании решений, действий

органов государственной власти, органов местного самоуправления, должностных лиц, государственных или муниципальных служащих А, в результате которых, по мнению указанных лиц, были нарушены их права и свободы или созданы препятствия к осуществлению ими прав и свобод либо на них незаконно возложена какая-либо обязанность или они незаконно привлечены к ответственности.

О: решений

В: Какие решения можно оспорить?

В соответствии со ст. 46 Конституции РФ и гл. 24 АПК РФ граждане и организации вправе обратиться в суд за защитой своих прав и свобод с заявлением об оспаривании решений А, действий органов государственной власти, органов местного самоуправления, должностных лиц, государственных или муниципальных служащих, в результате которых, по мнению указанных лиц, были нарушены их права и свободы или созданы препятствия к осуществлению ими прав и свобод либо на них незаконно возложена какая-либо обязанность или они незаконно привлечены к ответственности.

Однако, перефразируя тот же вопрос, мы получаем другой результат.

Когда вопрос формулируется более абстрактно, последовательность слов в предложении начинает иметь гораздо большее значение, чем его синтаксическая структура, что не соответствует смыслу предложения.

В рамках второго теста мы попытались определить субъектный состав и объект договора из более простого предложения: «*Между ООО «Кротвест и ООО «МедКо» заключен договор аренды в отношении нежилого помещения».*

О: АО «Кротвест» и ООО «МедоКо»

В: Каков субъект договора?

Между **АО «Кротвест» и ООО «МедоКо»** заключен договор аренды в отношении нежилого помещения.

О: нежилого помещения

В: Каков объект договора?

Между АО «Кротвест» и ООО «МедоКо» заключен договор аренды в отношении **нежилого помещения**.

Как видно из результатов, система DeepPavlov успешно определила объект и субъектный состав договора из простого предложения.

Усложнение задачи и изменение последовательности слов в предложении на входе снова демонстрирует жёсткую привязку семантического анализа к синтаксису и последовательности слов.

О: АО «Кротвест»

В: Каков субъект договора?

**АО «Кротвест»** заключило договор аренды в отношении нежилого помещения с ООО «МедоКо».

Субъект определяется корректно только том случае, когда стороны записаны подряд и напрямую связаны синтаксически (в данном случае предлогом «между» и соединительным союзом «и»). Стоит отметить, что с точки зрения юридической логики оба предложения содержат абсолютно идентичные смыслы: субъектами являются обе компании.

Однако в зависимости от формулировки вопроса ответ определяется неустойчиво.

О: АО «Кротвест»

В: Кто заключил договор?

**АО «Кротвест»** заключило договор аренды в отношении нежилого помещения с ООО «МедоКо».

О: с ООО «МедоКо»

В: Какими сторонами заключен договор?

АО «Кротвест» заключило договор аренды в отношении нежилого помещения с ООО «МедоКо» .

Для третьего теста мы проверили возможность вычленять факты, используя конверсивы (слова, выражающие отношения к одну и тому же событию с разных углов зрения). За основу было взято предложение с именованными сущностями (организация и лицо) и глаголом «продать». Модель хорошо идентифицировала вопросы, лексически дублирующие текст предложения.

О: ООО "Лесник"

В: Кто продал слэбы?

ООО "Лесник" продало Друнову В.С. партию деревянных слэбов.

О: Друнову В.С.

В: Кому продали партию?

ООО "Лесник" продало Друнову В.С. партию деревянных слэбов.

Когда же в вопросе применялся конверсив «купить», ответ снова привязывался к синтаксической зависимости вопроса и предложения, а не к семантическому концепту в целом.

О: ООО "Лесник"

В: Кто купил слэбы?

ООО "Лесник" продало Друнову В.С. партию деревянных слэбов.

Аналогичные операции были проделаны на более простом предложении: «Зоя продала Лене коляску».

О: Лене

В: Кому продали коляску?

Зоя продала Лене коляску.

О: Зоя

В: Кто продал коляску?

**Зоя** **A** продала Лене коляску.

О: Зоя продала Лене

В: Кто купил коляску?

**Зоя продала Лене** **A** коляску.

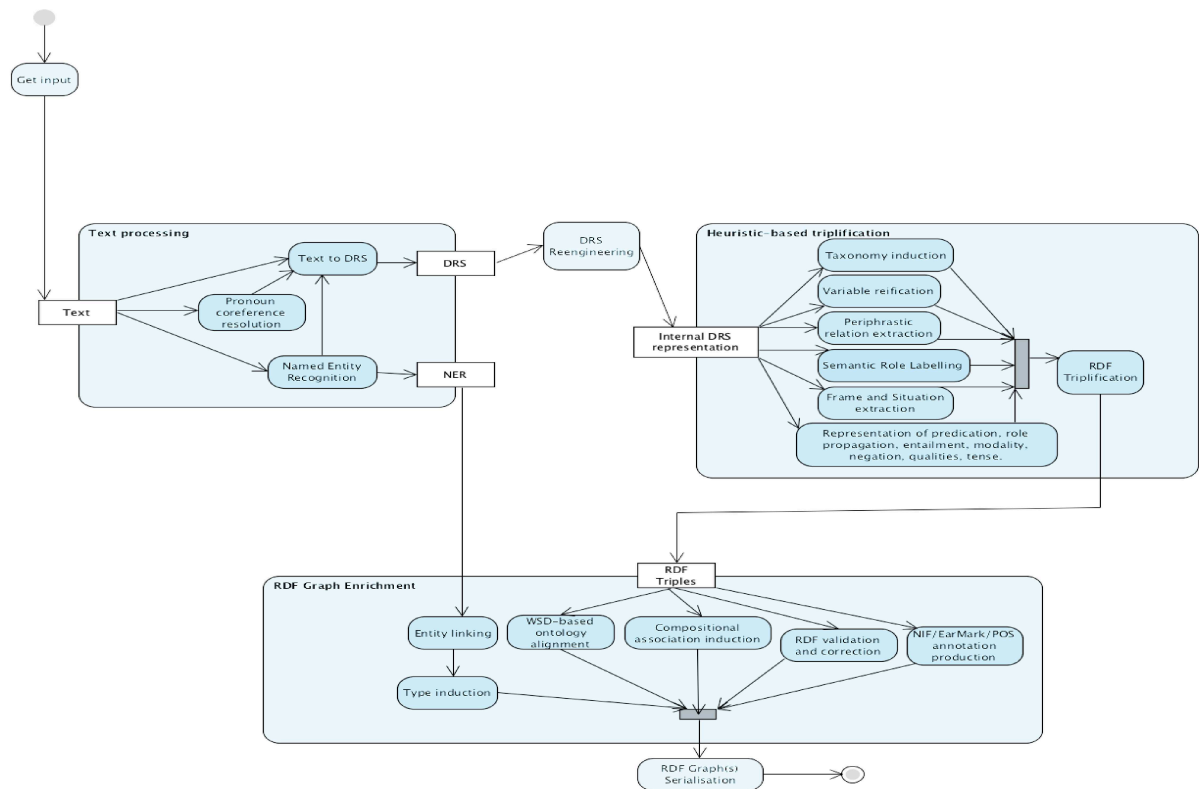
В данном тесте модель успешно идентифицировала как объект, так и субъектный состав отношений.

Помимо приведённых примеров мы провели множество других тестов QA-модели DeerPavlov, в том числе на предложениях, не связанных с юриспруденцией. В результате можно сказать, что модель показывает хорошие результаты при постановке вопросов, синтаксически и лексически близких к предложению на входе и, потенциально, может быть использована в качестве поисковой системы для документов большого объёма.

Однако для качественного контекстуального анализа и, как следствие, выделения точных данных и значимых смыслов из текста с помощью данной модели видится невозможным. А ведь именно эти задачи представляют особую важность для инструментов, применяемых в юриспруденции.

### 3.2.1.2. FRED machine reader for the Semantic Web

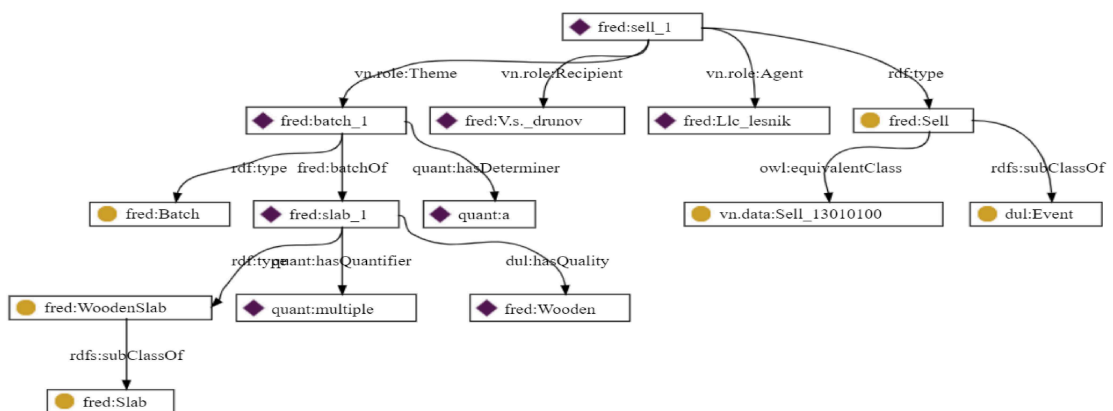
Ещё одним инструментом семантической обработки текста, привлёкшим наше внимание, является семантический парсер FRED. Архитектура FRED построена на применении нейронных сетей, принципов фреймовой семантики и онтологического подхода, в результате чего анализируемый текст систематизируется путём выстраивания семантических связей между элементами и множеством существующих онтологий.



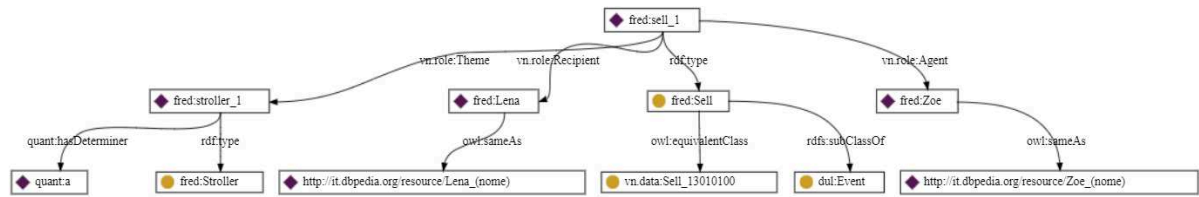
Система успешно выделяет семантические группы и предоставляет ссылки на используемые онтологии, что позволяет продолжить процессинг текста через онтологические отношения. Предлагаем более детально рассмотреть достоинства и недостатки данного подхода на примерах.

В качестве исходных данных для первого теста мы выбрали следующие простые предложения:

*«ООО «Лесник» продало Друнову В.С. партию деревянных слэбов».*



*«Зоя продала Лене коляску».*



Оба примера связаны с отношениями купли-продажи объекта.

С точки зрения гражданского права РФ договор купли-продажи включает в себя сторон (продавец и покупатель), объект (объект гражданского оборота, отчуждаемый продавцом в собственность покупателя) и предмет (действия, совершаемые сторонами для достижения желаемого правового результата, а именно – продавец обязуется передать имущество в собственность покупателя, а покупатель – принять объект и оплатить его стоимость).

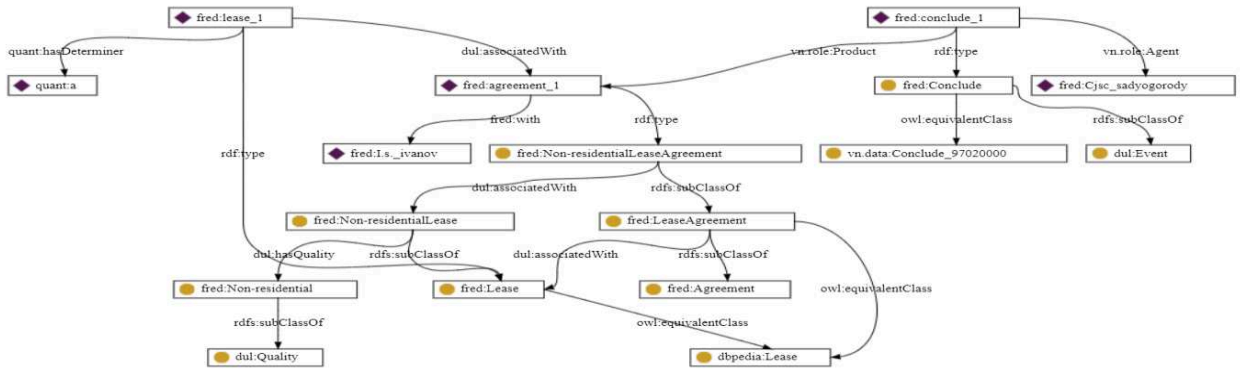
Рассмотрим, как факты, связанные с продажей, структурируются с помощью FRED:

- выявлено событие «Продажа» (в графе оно не предполагает обратного события – «Покупка», однако такая возможность должна подразумеваться отношениями в онтологиях);
- событие имеет Агента (продавца) и Реципиента (покупателя);
- событие имеет «Тему» продажи (продукт) с его атрибутами, иными словами – материальный объект отношений.

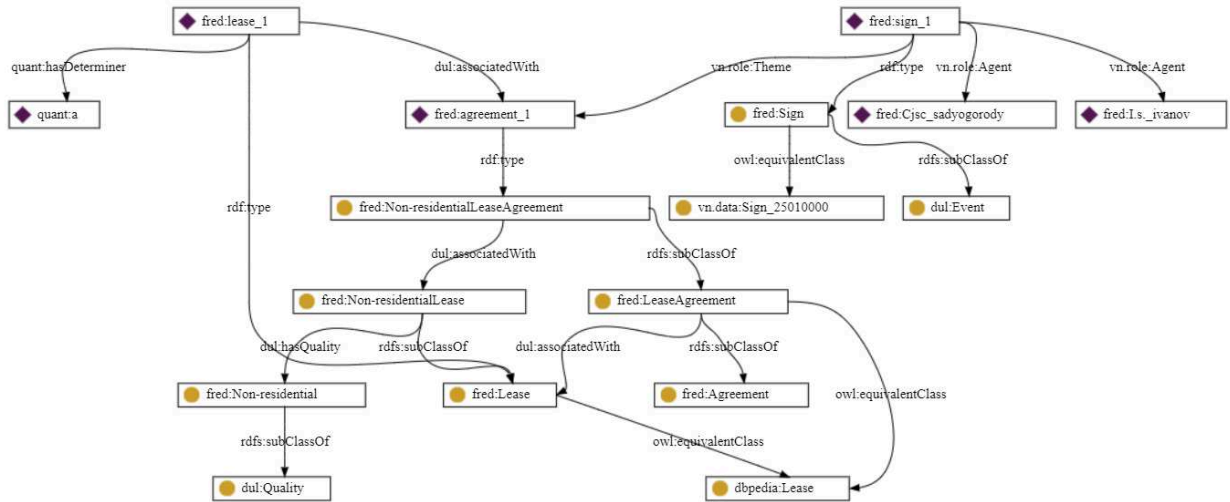
Стоит отметить, что имена физических лиц и наименований организаций во FRED определяются так же неустойчиво (сравните связи агента и реципиента в первом и втором предложении): если во втором предложении «Зоя» и «Лена» соотносились с соответствующими именами в онтологии (иначе говоря, были распознаны в качестве имён), то в первом предложении Агент «ООО Лесник» и Реципиент «Друнов В.С.» не соотносятся с классами онтологии вовсе, что говорит о неидентификации наименований организаций в качестве таковых.

Далее рассмотрим пример с синтаксической близостью сторон в предложениях о заключении договора на следующих примерах:

*«ЗАО СадыОгороды» заключило договор аренды нежилого помещения с И.С. Ивановым».*



«ЗАО СадыОгороды» и И.С. Иванов заключили договор аренды нежилого помещения».



В первом примере на графе событие имеет только одну сторону (в качестве Агента идентифицировано «ЗАО СадыОгороды»).

Второй же фактический Агент данного предложения «И.С. Иванов» ошибочно ассоциирован с событием через «Продукт» (договор) связью "с" – «Договор "с" И.С. Ивановым».

С точки зрения семантики и юридической логики эта связь уместнее в случае «Договор «с» правками (приложением; доп. соглашением и т.д.)».

Во втором же примере на графе, верно, выделяются обе стороны договора («ЗАО Сады Огороды» «И.С. Иванов» имеют роль Агента).

Данные тесты наглядно демонстрируют наличие общей проблемы у FRED и DeepPavlov при семантическом анализе, связанной с сильной привязкой к синтаксису и последовательности слов в предложениях. FRED в сравнении с

DeepPavlov, на первый взгляд, видится инструментом, позволяющим выделять факты и семантические связи между ними более конкретно. Однако, при детальном рассмотрении у данного подхода обнаруживаются те же проблемы, что и у DeepPavlov.

### 3.2.2. Named Entity Recognition (NER)

Ещё один инструмент, который широко применяется в компьютерной лингвистике, – Named Entity Recognition (NER). Инструменты NER позволяют распознавать в тексте устойчивые именные сущности по типу таких структурированных данных: дата и время, суммы и числовые величины, адреса, наименования географических объектов, регистрационные и идентификационные номера, ФИО и наименования компаний и др., а также определять их принадлежность к той или иной группе или категории.

На тему подбора оптимальных для русского языка инструментов NER и библиотек уже написано некоторое количество статей. Вкратце стоит отметить, что на сегодняшний день существует множество хороших моделей, пока дело не доходит до русского языка...

Здесь мы сталкиваемся с двумя основными проблемами: как правило, ограниченным количеством классов и/или rule-based системами, имеющими в перспективе некоторые ограничения в развитии и, как следствие, риски в их применении. Но готовых решений нет и, в сущности, не может быть без адаптации инструмента к предметной области и решаемым задачам.

Для применения NER в юридической сфере требуется высокая степень детализации и точности разметки ФИО и наименований организаций, классов документов, чисел, торговых марок, наименования объектов гражданского оборота и других сущностей.

По своей сути классификация именованных сущностей должна быть построена юристами с учётом действующего правового режима.

С точки зрения профессиональной логики такой классификатор должен включать в себя исчерпывающий и закрытый перечень участников оборота (физическое лицо, индивидуальный предприниматель, юридическое лицо с учётом всех организационно-правовых форм), объектов (недвижимое и движимое имущество с учётом всех разновидностей и др.) и должен быть сформирован по принципу – совокупность дочерних сущностей образует родительскую.

Например, помещение, здание, земельный участок, морское судно и др. в совокупности образует категорию «недвижимое имущество», а недвижимое имущество в совокупности с движимым есть «вещь».

На практике же, существующие инструменты распознавания именованных сущностей содержат в себе классификаторы, адаптированные под универсальное применение без привязки к конкретной области знаний. Например, библиотека DeepPavlov предлагает следующие типы распознаваемых сущностей.

PERSON	People including fictional
NORP	Nationalities or religious or political groups
FACILITY	Buildings, airports, highways, bridges, etc.
ORGANIZATION	Companies, agencies, institutions, etc.
GPE	Countries, cities, states
LOCATION	Non-GPE locations, mountain ranges, bodies of water
PRODUCT	Vehicles, weapons, foods, etc. (Not services)
EVENT	Named hurricanes, battles, wars, sports events, etc.
WORK OF ART	Titles of books, songs, etc.
LAW	Named documents made into laws
LANGUAGE	Any named language
DATE	Absolute or relative dates or periods
TIME	Times smaller than a day
PERCENT	Percentage (including "%")
MONEY	Monetary values, including unit
QUANTITY	Measurements, as of weight or distance
ORDINAL	"first", "second"
CARDINAL	Numerals that do not fall under another type

На первый взгляд, такая классификация представляется вполне разумной, так как во многом совпадает с логикой/форматом восприятия обычного человека.

Однако если посмотреть на такую структуру глазами юриста-профессионала, то обнаруживается ряд существенных проблем:

- предлагается одноуровневая структура, которая не укладывается в рамки российского законодательства: в одну категорию «Facility» объединены все объекты материального мира, являющиеся по своей сути недвижимостью (здания, мосты и др.), а в «Product» – движимое имущество, не учитывая, что все эти объекты являются вещами;
- самостоятельной группой выступает «Work of art», включающая в себя произведения искусства (с точки зрения права – результаты интеллектуальной деятельности) и др.

Используемая в DeepPavlov классификация хорошо подходит для общих целей и воспроизводит некоторую верхнеуровневую группировку объектов и субъектов материального мира. Но такой подход полностью игнорирует

фундаментальные основы российского права: особенности объектов гражданского оборота, субъекты правоотношений, их правовой статус и др.

Книга в материальном понимании – это вещь, аналогично велосипеду, но которая одновременно является и произведением искусства, охраняемым как результат интеллектуальной деятельности.

По большому счёту, этот факт говорит о необходимости участия экспертов предметной области в построении классификатора «с нуля» в соответствии юридической базой знаний.

При этом юристы имеют возможность фактически создать профессиональную таксономию (экспертное видение предметной области) под актуальные задачи.

Важность корректного распознавания именованных сущностей заключается в том, что именованные сущности являются фундаментом для семантического анализа текстов.

При этом верно и обратное: семантический анализ способствует более точной разметке именованных сущностей, когда только из контекста возможно вычленить искомые атрибуты.

Достаточно распространённый пример, иллюстрирующий необходимость использования глубокого контекстуального анализа, – эллипсис (пропуск элемента высказывания, легко восстанавливаемого в данном контексте), появляющийся в текстах ближе к середине-концу изложения.

Например, тот случай, когда в первом абзаце говорится об «ИП Иванов И.С.», который должен быть идентифицирован в качестве индивидуального предпринимателя, а далее повествование содержит различные вариации той же сущности, в том числе «Иванов», который без контекстуального анализа будет размечен в качестве физического лица).

### **3.2.3. Синтаксический парсинг**

Третьим важным инструментом процессинга текста выступает синтаксический анализ, выполняемый с помощью различных синтаксических парсеров. Синтаксический парсер представляет собой инструмент анализа предложений на основе его синтаксической структуры и представления данных в виде дерева зависимостей, выстроенного между словами.

Выбор синтаксического парсера, очевидно, определяет работоспособность семантического анализатора, поскольку синтаксические связи (в т.ч. их тип и место) напрямую определяют результат семантической обработки.

Для исследования работоспособности синтаксических анализаторов на текстах юридического характера, имеющих структурную специфику (юридический стиль характеризуется инкорпорированием в сжатое повествование формальной информации типа наименования документов, паспортных данных и пр., аббревиатур, насыщенностью оборотами, приложениями, подчинительными и сочинительными связями, пояснениями и др.), нами была создана тестовая коллекция предложений разной степени сложности, взятых из реальных юридических документов.

Предложения коллекции были дифференцированы и разбиты на три группы: простые предложения, предложения средней и высокой степени сложности. Для понимания сложности задачи приведём примеры каждой из групп предложений из тестовой выборки:

1. Простое предложение: *«Между Истцом и Ответчиком заключён Договор аренды от 01.08.2012 г. в отношении нежилого помещения».*

2. Предложение средней степени сложности: *«Принимая во внимание вышеизложенное, Договор купли-продажи доли является ничтожным в силу статьи, поэтому не влечёт юридических последствий за исключением тех, которые связаны с его недействительностью и недействителен с момента его совершения».*

3. Предложение высокой степени сложности: *«Принимая во внимание вышеизложенное, Договор купли-продажи доли в размере 75% уставного капитала Общества от 26.10.2006 г., заключённый между Компанией Марс Систем и Закрытым акционерным обществом «Консалтинговая фирма «СТН МРТ», является ничтожным в силу ст. 168 ГК РФ, поскольку заключён с нарушением п.3 ст. 154 ГК РФ, поэтому не влечёт юридических последствий за исключением тех, которые связаны с его недействительностью и недействителен с момента его совершения».*

Мы проанализировали большинство доступных парсеров (ниже приведена основная их часть) на данной коллекции юридических предложений разной степени сложности и пришли к выводу, что качественный синтаксический парсинг на основе существующих инструментов возможен только для простых предложений и предложений средней степени сложности (с рядом оговорок).

Парсинг предложений высокой степени сложности пока недоступен и требует существенных доработок.

При тестировании парсинга предложений средней степени сложности был выбран следующий фрагмент: *«В разъяснениях, данных в п. 16 Постановления пленума ВС РФ и Пленума ВАС РФ «О некоторых вопросах применения*

*Федерального закона», указано на то, что выход участника из общества осуществляется на основании его заявления».*

Подробности тестов различных синтаксических парсеров приведены ниже.

### 3.2.3.1. АОТ

Проект «Автоматическая обработка текстов» (АОТ) приостановил своё развитие несколько лет назад. Как заявляли его создатели, подход, используемый в АОТ, скорее можно назвать консервативным, чем революционным, поскольку общая идея, объясняющая сущность естественного языка, отсутствует.

Максимально приблизить человеческий язык к современному компьютеру возможно только при помощи грамотной декомпиляции языковых механизмов.

Тем не менее для нас было важно апробировать данную систему для сравнения её с грамматикой зависимостей, поскольку АОТ применяет грамматику непосредственно составляющих.



На скриншоте с результатами теста заметно, что не все слова и сокращения («ВАС РФ», «указано») в предложении имеют связи со словами и группами слов. «ВАС» и «РФ» должны быть связаны по типу «генит\_иг», а «указано» должно объединиться с предложной группой «на то...» связью «кр.прч.».

Помимо этого, большое количество именных, предложных и прочих групп не связаны друг с другом. К примеру, группа «выход участника» синтаксически связана с группой «из общества», что не следует из выведенного системой разбора.

Дело в том, что грамматика непосредственно составляющих с трудом перекладывается на структуру русского языка с его нефиксированной последовательностью слов в предложении, и, соответственно, предложение теряет значительное количество синтаксических связей, в том числе влияющих на смысл.

Отсюда следует, что главным минусом данного подхода является неполнота связей. Иначе говоря, некоторые токены остаются «в воздухе», что недопустимо при создании графа и дальнейшем семантическом анализе.



Также система дробит числовые показатели и, следовательно, находит новые несуществующие с точки зрения смысла связи.

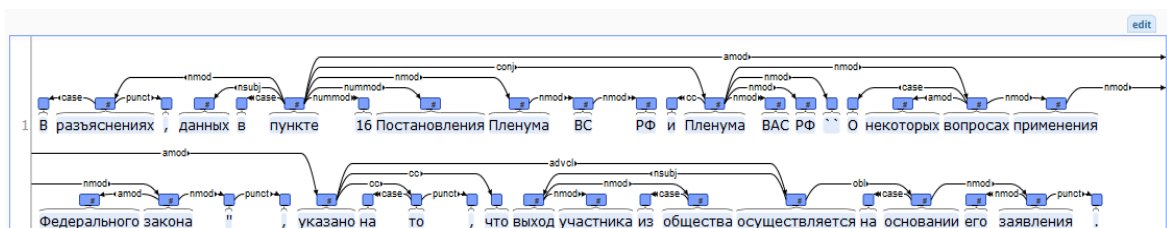
К преимуществам ЭТАП можно отнести присутствие дополнительных лексических функций/связей, позволяющих более глубоко работать с семантикой на следующих этапах

### 3.2.3.3. Stanford

Синтаксический парсер Stanford – парсер, разработанный в Стендфордском университете. Данный парсер использует модель arc-standard system, где выбор действия осуществляется с помощью нейронной сети. Особенностью данного парсера является то, что он изначально был создан и обучен для применения на англоязычных текстах, при этом грамматика английского языка предполагает только 15 типов связей.

При адаптации парсера для русского языка типы связей были сохранены. Однако русский язык содержит порядка 30 типов связей, что порождает значительные трудности в использовании данного парсера на русскоязычных текстах.

В процессе тестирования Stanford выдаёт минимальное количество разнородных, но зачастую несущественных ошибок, верно выявляя структуры сложносочинённых и сложноподчинённых предложений.



Здесь сочинительной связью «conj» система объединила категории «пункте» и «Пленума», хотя из предыдущего разбора нам уже известно, что эта связь закреплена за словами «Пленума» и «Пленума».

Стоит также обратить внимание на следующую особенность: при том что словосочетание «Пленума ВС РФ» с его аббревиатурами разобрано верно (последовательная зависимость «nmod»: от «Пленума» к «ВС», от «ВС» к «РФ»), аналогичное словосочетание «Пленума ВАС РФ» уже требует корректив (вместо последовательности от «Пленума» как к «ВАС», так и к «РФ», необходима

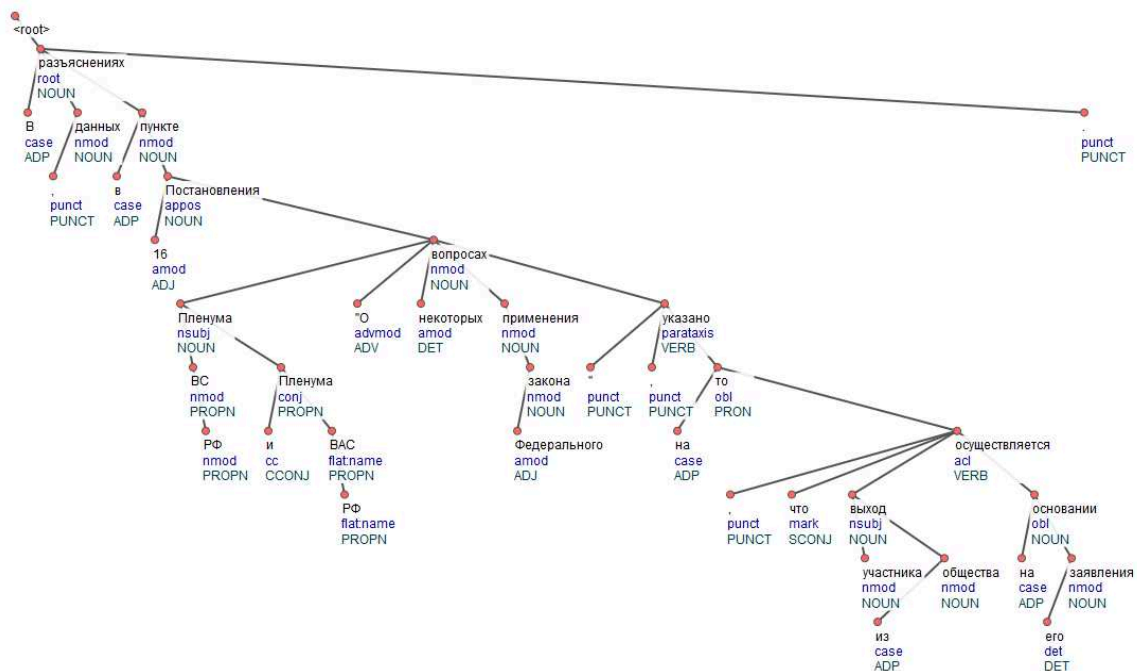
последовательная связь, аналогичная предыдущему словосочетанию), что говорит о нестабильном результате вывода. При этом относительно качественно разобраны структурные части предложения (причастные обороты, подчинительная часть).

В итоге к преимуществам Stanford мы можем отнести устойчивую работу при анализе структуры предложения (что является существенным аргументом в пользу парсера) и полный разбор предложения без пропусков. К недостаткам – скорость обработки, некоторая хаотичность выделения связей между токенами (от глагола к прилагательному, а уже после через него к существительному).

### 3.2.3.4. UDPipe 2.4

Модель Universal Dependencies предобучена на нескольких русских размеченных корпусах, среди которых к сравнительному анализу были взяты следующие: GSD, taiga и SynTagRus. Безусловно, самой качественной среди них оказалась модель на SynTagRus – крупнейшем русском размеченном корпусе на сегодня.

#### UDP GSD



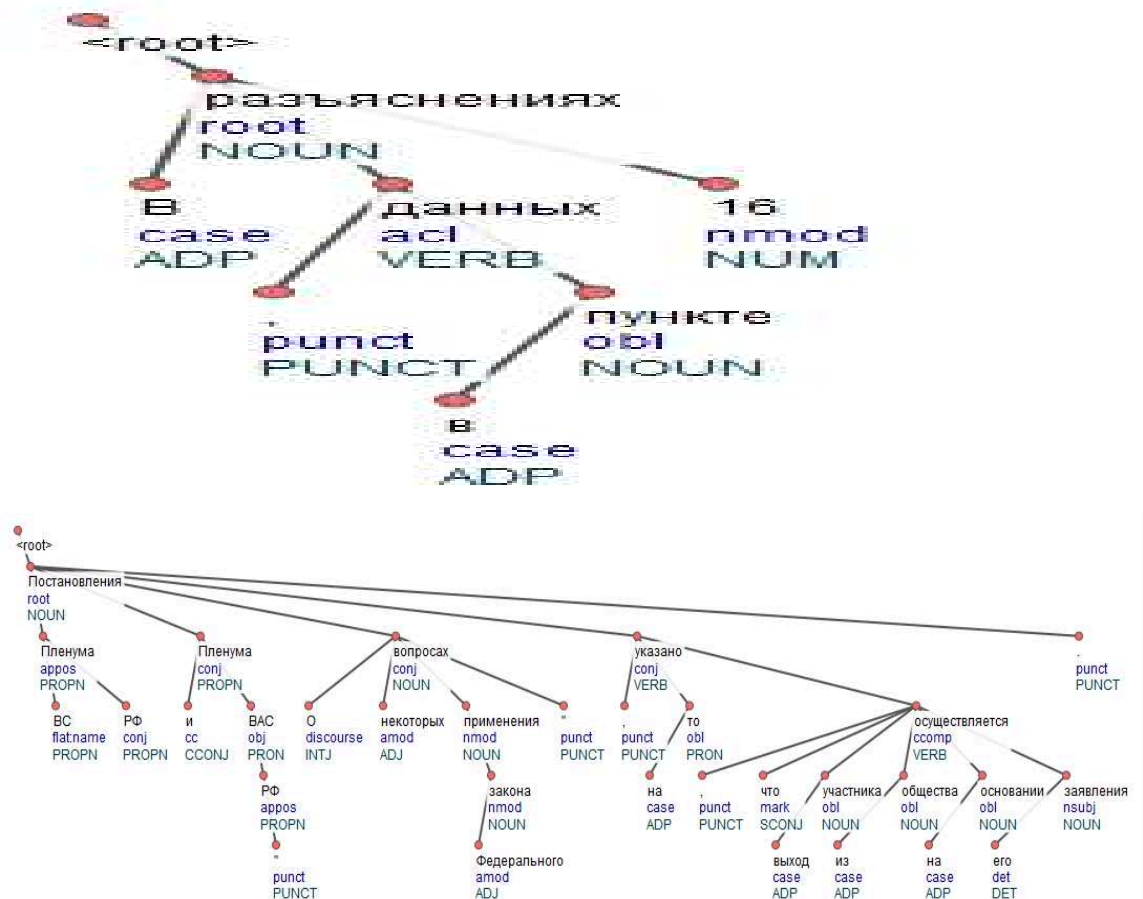
UDPipe, обученный на GSD, выдаёт частые ошибки внутри структуры предложений. В данном дереве имеющие прямую синтаксическую связь слова

«разъяснениях» и «указано», разделены друг от друга 4 уровнями связей. Кроме того за root принимаются не только глаголы, но и существительные (не всегда подлежащие).

На примере видно ошибочное определение «root» (разъяснениях), что значительно исказит дальнейший семантический анализ.

Также часто и бессистемно дроблению подвергаются длинные предложения. Насыщенность однородными членами также способна сломать структуру дерева. Зачастую, эти особенности обнаруживается на предложениях средней и высокой степени сложности.

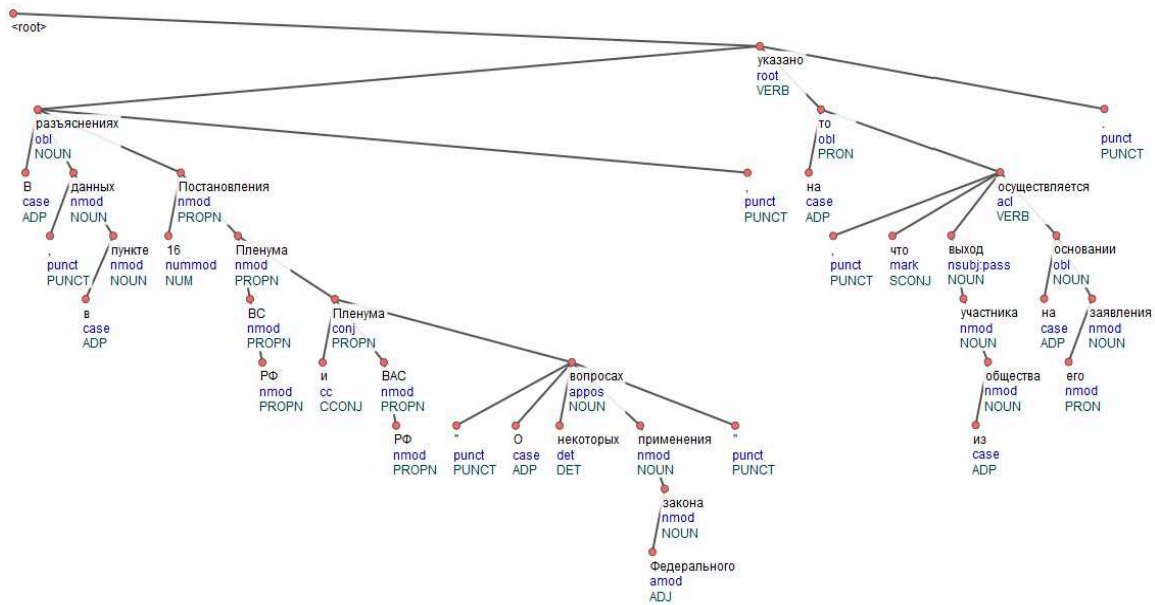
## UDP taiga



На аналогичном примере UDP 2.4 taiga продемонстрировала огромное количество искажений, но главной принципиальной ошибкой стало деление одного предложения на два дерева.

UDPipe, обученный на Taiga даёт аналогичные предыдущему корпусу ошибки, однако дробление предложения на несколько деревьев возникает зачастую уже на предложениях средней сложности.

## UDP SynTagRus



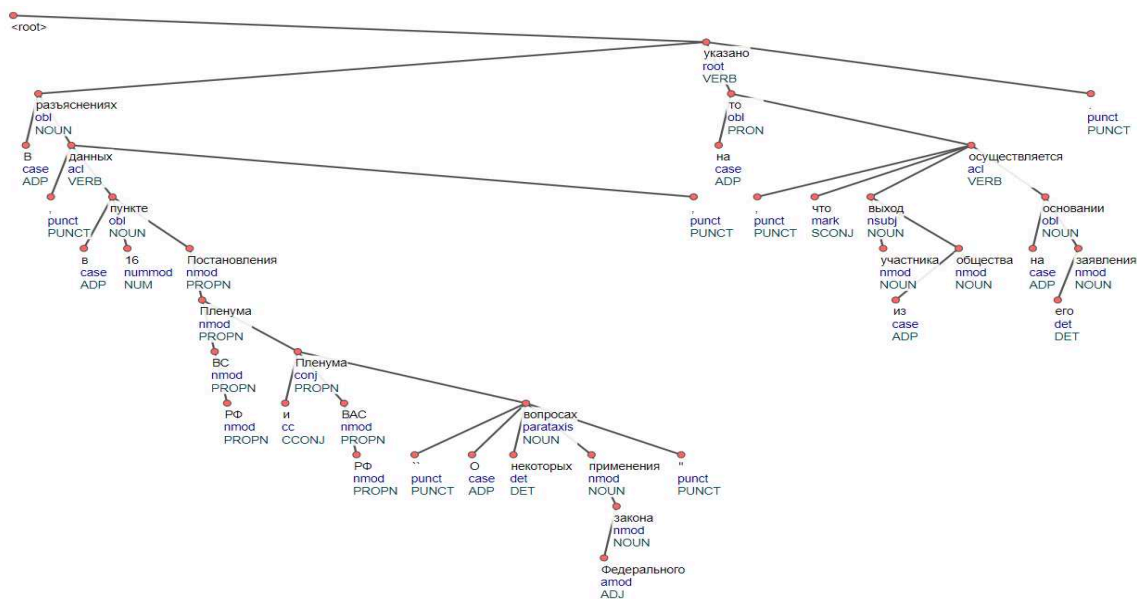
Иначе обстоят дела с UDP 2.4 SynTagRus. Он вывел качественно проанализированное дерево, недочётами которого стала последовательность связей (в словосочетании «выход участника из общества» «выход» должен быть напрямую связан с «обществом», а не через связь с «участником») и обработка числовых показателей (очевидно, «16»-ым является «пункт», а отнюдь не «Постановление»).

UDPipe на SynTagRus имеет более стойкие, отслеживаемые ошибки в предложениях высокой степени сложности. Однако структура остаётся сохранна, за исключением случаев особо нагромождённых предложений, которые парсер имеет свойство дробить.

В целом к явным плюсам UDPipe можно отнести скорость и обновляемость.

### 3.2.3.5. DeepPavlov

Последней была протестирована модель DeepPavlov, предобученная на корпусе SynTagRus с помощью BERT.



Модель показала очень достойные результаты, в целом приближенные к UDPipe 2.4 на SynTagRus. Наблюдаются проблемы на сложных многоуровневых предложениях, ошибки в месте связи или в её типе.

На данном примере видим, пожалуй, единственный неочевидный нюанс с «Постановлением... «О вопросах...», где «Постановление» и «вопросах» должны иметь не опосредованную, а прямую связь (в контексте предложения наименование «О вопросах...» присвоено не Пленуму, а Постановлению).

### 3.2.3.6. Итоговые результаты тестирования синтаксических парсеров

Результаты тестирования существующих синтаксических парсеров можно представить в следующей таблице (инструменты ранжированы в зависимости от результатов – от наиболее успешного к наименее).

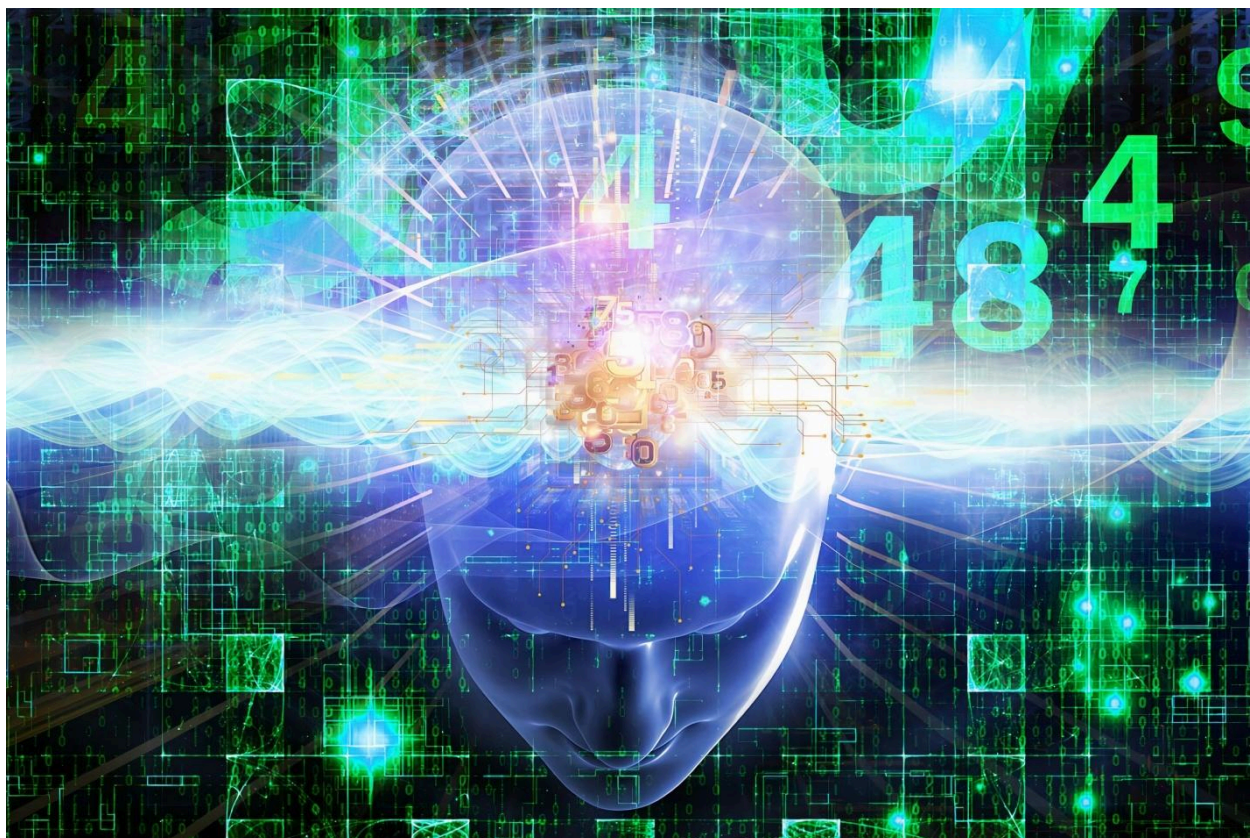
Наименование	Преимущества	Недостатки
DeepPavlov	скорость, обновляемость и сохранение структуры предложения	нарушение последовательности связей в сложных предложениях

Наименование	Преимущества	Недостатки
UDPipe 2.4 SynTagRus	скорость, обновляемость и сохранение структуры предложения	нарушение последовательности связей в сложных предложениях и числовых значениях
Stanford	устойчивая работа при анализе структуры предложения и полный разбор предложения без пропусков	низкая скорость обработки и хаотичность выделения связей между токенами
ЭТАП	присутствие дополнительных лексических функций/связей, позволяющих более глубоко работать с семантикой на следующих этапах	нарушение семантической логики связей, дробление числовых показателей и нахождение несуществующих связей
UDPipe 2.4 GSD	-	дробление предложений на два и более деревьев, ошибки в последовательности и типах связей
UDPipe 2.4 Taiga	-	дробление предложений на два и более деревьев
АОТ	-	неполнота связей (некоторые токены остаются "в воздухе", что недопустимо при создании графа и дальнейшем семантическом анализе)

Исследованные инструменты показали различные результаты. Некоторые парсеры, например DeepPavlov, UDPipe 2.4 SinTagRus и Stanford, продемонстрировали относительно хороший уровень разбора простых предложений и предложений средней степени сложности, однако общим для них является неэффективность на юридических текстах. Для достижения приемлемого уровня

разбора текстов правовых документов данные модели подлежат значительной доработке.

## Заключение



### 4.1. Выводы по итогам тестирования

Результаты исследования существующих решений в области процессинга русскоязычного текста привели нас к выводу, что представленные на рынке инструменты имеют универсальный характер и неприменимы в существующем виде для достижения практических результатов в анализе слабоструктурированных и неструктурированных правовых документов. Причин тому несколько.

Основная проблема, присущая всем представленным решениям, заключается в том, что продукт создан не экспертами предметной области, в которой он применяется.

Идея создания инструментов автоматизации юридической работы без участия юристов высокой квалификации изначально обречена на неудачу, поскольку без понимая терминологии, её значений и классификаций, а также самых глубинных взаимосвязей невозможно воссоздать «юридическую картину мира».

Во многом данная ситуация связана с тем, что на рынке доминируют подход, при котором идеологами проектов по созданию решений автоматизации выступают IT-разработчики и специалисты в области data science, которые не знакомы на должном уровне с особенностями юридического мышления и не погружены в реальную практику, в которой может применяться то или иное решение.

Кроме того, при создании программных продуктов многими разработчиками преследуется логичная цель – максимально широкая интеграция ПО в различные сферы.

Наиболее простой подход для этого – создание универсальной платформы, внедрение которой в новую предметную область потребует незначительных доработок.

Данный подход имеет свои преимущества и недостатки, но в случае создания инструментов автоматизации работы юриста он неприменим.

Для использования таких продуктов конечные пользователи (практикующие юристы) вынуждены подстраиваться и адаптироваться под их условия и особенности, а в данной сфере должно быть наоборот – продукт изначально должен создаваться исходя из потребностей и задач пользователей. Только такой подход позволит добиться качественных результатов.

Третьим фактором, который до настоящего времени не позволил реализовать в полном объёме процессинг русскоязычных текстов на достаточном уровне, является недостаточность финансирования научных разработок и отсутствие государственных и частных инвестиций в данную область знаний. Существующие научные группы и институты благодаря собственной воле и энтузиазму достигли хороших практических результатов в создании инструментов обработки текста, однако ограниченность бюджета не позволяет им продвинуться дальше. Их зарубежные коллеги достигли гораздо больших результатов благодаря активному участию государства, финансирующего подобные проекты.

По нашему мнению, государственная поддержка исследователей и проектов в области NLP и искусственного интеллекта позволит совершить настоящий прорыв в краткосрочной перспективе.

## 4.2. Дальнейшее развитие

Как показывает практика, создать высокоэффективное программное решение, которое может быть интегрировано в конкретную предметную область, невозможно без участия экспертов из данной области, профессиональный опыт и логика которых ложатся в основу машинных алгоритмов. Именно поэтому мы считаем, что создание «цифрового юриста» (юридического ИИ) и содержательная автоматизация юридической функции:

- возможно только в результате глубинного погружения в предметную область;
- находятся на пересечении 3 различных областей знаний: юриспруденция, лингвистика и IT.

В результате тесного взаимодействия специалистов из этих областей будут созданы новые уникальные для рынка компетенции.

Данные компетенции находятся на стыке нескольких областей – юриспруденции, лингвистики, программирования и инженерии знаний, что приведёт к формированию принципиально новых профессий, отсутствие которых сегодня является одним из наиболее существенных факторов, сдерживающих развитие рынка технологий искусственного интеллекта в России.

Решение задачи процессинга неструктурированного русскоязычного текста в области юриспруденции требует иного подхода к использованию инструментов NLP: они должны учитывать на фундаментальном уровне юридические концепты и базироваться на графах знаний, которые изначально созданы для решения узкопрофильных юридических задач.

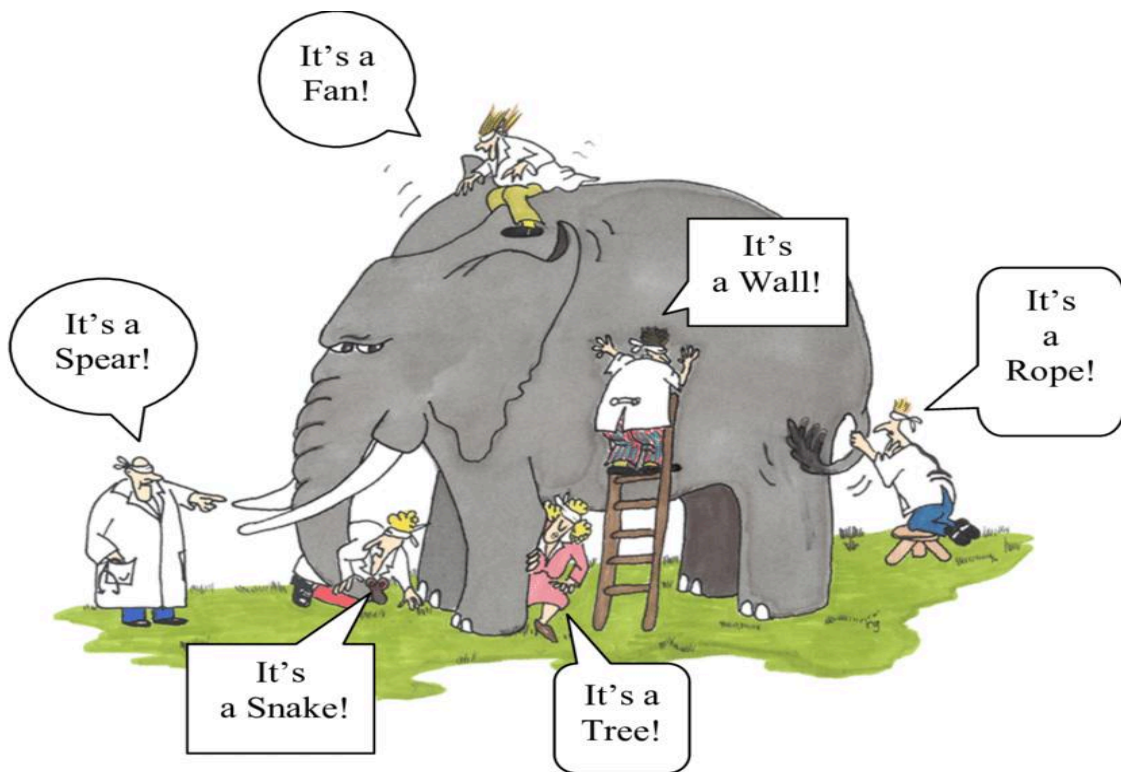
Это позволит воссоздать юридическую «картину мира» в цифровом формате и трансформировать юридическую логику в машинные алгоритмы. Безусловно, большое значение в данном вопросе имеет достаточный уровень финансирования и поддержки подобных проектов, которое возможно только при активном участии государства.

## Статья 2

### Введение

Анализ актуальной в настоящее время темы искусственного интеллекта и его применения в области юриспруденции мы начали с обзора инструментов синтаксического и семантического анализа текстов, которые применяются при разработке LegalTech-решений.

В комментариях к предыдущей статье красной нитью проходил очень непростой вопрос: а почему бы не извлекать из текста все имеющиеся в нем смыслы? В чем здесь сложность? Такой вопрос – крайне показателен, поэтому мы решили уделить ему более пристальное внимание и дать максимально развёрнутый ответ.



Итак, сегодня мы постараемся ответить на ряд, казалось бы, простых вопросов:

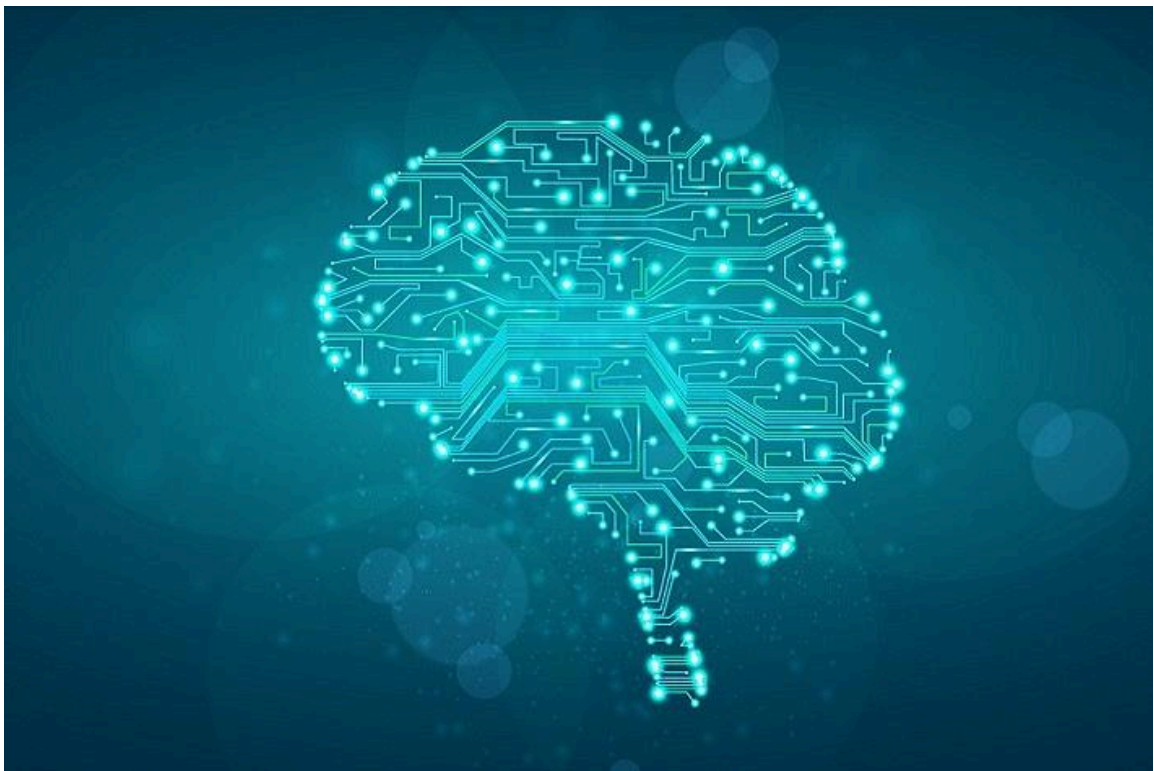
1. Как найти в тексте «все смыслы»? В чем различия в восприятии текста между экспертом в предметной области (юристом) и неспециалистом (разработчиком)?

2. Как эти различия влияют на разработку соответствующих IT-решений?

### 1. Legal AI: миф или реальность

В тексте нашей первой статьи мы определили основные задачи, которые необходимо решить на пути к созданию Legal Artificial Intelligence (далее по тексту – «Legal AI»). В качестве одной из ключевых и первостепенных задач мы выделили необходимость обучения машины смысловому пониманию документов на уровне, сравнимым с восприятием профессионального юриста, а равно — создание инструментов процессинга сложных юридических текстов на русском языке.

Мы детально исследовали на практике некоторые из доступных лингвистических инструментов обработки текста, но результаты и выводы, к которым мы пришли, получили неоднозначную оценку и вызвали дискуссию в профессиональном сообществе. Мы получили множество вопросов от специалистов, связанных с практической реализуемостью данных задач и перспектив применения технологий в этом направлении. Среди вопросов и комментариев, на наш взгляд, прослеживается некий пессимизм, отсутствие веры в технологии и их практическую применимость для решения поставленных задач, а также тезис о том, что Legal AI создать невозможно в силу различных причин и наличия трудных «нерешаемых» задач.



Мы не согласны с данной позицией. Текущего уровня развития технологий вполне достаточно, чтобы создать полноценный юридический искусственный интеллект, что и будет сделано в обозримом будущем.

Фундаментальные теоретические разработки продвинулись достаточно далеко, чтобы можно было перейти к их практической реализации. Очевидно, что частные проблемы и трудности в данной сфере, тормозящие темпы развития таких продуктов, действительно, существуют.

Но данные проблемы решаемы, однако пути решения требуют кропотливой и долгой пошаговой работы, которой многие опасаются.

Многие представители сферы LegalTech, сталкиваясь с трудностями и ограничениями технологий, отказываются от первоначальных концепций и создают довольно успешные пользовательские продукты с ограниченным функционалом.

Например, известный сервис DoNotPay, чат-бот, позволяющий обрабатывать однотипные юридические запросы граждан по решению бытовых споров и конфликтов: оспаривать штрафы за неправильную парковку, готовить иски и претензии о компенсации за задержку авиарейсов и поездов, а также требования о страховых выплатах.

Данный продукт, безусловно, является в некоторой степени прорывным, но он не решил главную задачу – создание полноценного «автоматизированного юриста», не привязанного к выборке типовых успешных кейсов. Этот недостаток ограничивает в том числе и перспективы развития сервиса. Руководствуясь количественным подходом, разработчики продукта расширяют набор типовых ситуаций, которые могут быть обработаны ботом, но все это упирается в обозначенное ограничение, а значит настанет момент, когда придётся либо вернуться к нерешённым проблемам, либо остановить развитие продукта.

## **1.1. Что такое Legal AI?**

Говоря о Legal AI, мы не занимаемся подменой понятий и не пытаемся поставить знак равно между искусственным интеллектом и машинным обучением / нейронными сетями. Для нас принципиально важной является именно автоматизация творческих задач, которые традиционно нельзя решить без участия человека-эксперта.

Поэтому инструменты Legal AI нацелены на замену человека, а не на оптимизацию решения отдельных задач в его деятельности.

Глобальная функциональная задача и цель Legal AI – трансформация юридической функции и создание полноценного цифрового эксперта, который способен анализировать данные и генерировать выводы на основе юридической логики, заменив тем самым команду высоко специализированных юристов.

При успешной реализации данных задач общество получит инструмент, способный отвечать на сложные вопросы предметной области, например, «Кто может подписать договор?» или «Какая ответственность может наступить в случае превышения полномочий?».

Помимо этого, Legal AI является инструментом, выходящим за рамки привычного понимания классической юриспруденции, а перспективы его практического применения не ограничиваются анализом рисков и обработкой типичных юридических документов (договоров, исковых заявлений и др.). Любой письменный документ содержит в себе юридически значимые сущности, которые могут быть связаны с теми или иными правовыми последствиями для организации или третьих лиц.

Например, с момента поступления и регистрации письма, содержащего обращение гражданина или организации в любой государственный орган, начинает идти срок, предусмотренный законом или иным нормативным актом для его рассмотрения.

Таким образом, область Legal AI охватывается весь документооборот в деятельности любой компании или государственного органа.

Один из многочисленных примеров – закупочные процедуры по выбору поставщика товаров/услуг/работ. Каждая из стадий закупочной процедуры (публикация сведений и конкурсной документации, получение заявок, принятие решений, публикация итогов и взаимодействие с участниками конкурсного отбора) предполагает характерные для нее документы, действия и связи между субъектами, которые можно формализовать в виде конечного набора правил, требований и, как следствие, – можно автоматизировать с помощью Legal AI.

Мы считаем, что создание и практическое применение Legal AI в обозначенных областях возможно, но только при полном осознании глубины существующих проблем и правильном подходе к созданию продукта, основанном на необходимости тесного взаимодействия со специалистами предметной области и создания онтологий и графов знаний, воспроизводящих логику юриста-профессионала.

## 2. Новый подход к созданию продукта

### 2.1. Построение диалога между экспертами

Взаимодействие команды и правильный подход к созданию конечного продукта является важнейшей составляющей успеха. Мы неоднократно отмечали в качестве оптимального подхода для создания LegalTech-продуктов необходимость тесного взаимодействия:

- юристов, являющихся носителями глубоких экспертных знаний в предметной области;
- лингвистов, разрабатывающих методики и инструменты процессинга текста;
- разработчиков, осуществляющих практическое воплощение продукта.

Однако построение такого взаимодействия сопряжено с рядом сложностей, одной из которых является проблема в коммуникации между разработчиками и юристами, что в результате может привести к несогласованности командной работы и недостижению необходимого результата.

Юрист и разработчик говорят на разных языках и чаще всего не понимают друг друга, поскольку это классическое столкновение противоположных форматов мышления. Мышление разработчика основано на применении дискретной логики, которая проявляется в том числе в подходе, при котором все процессы могут быть алгоритмизированы тем или иным образом.

Мышление юриста базируется на противоположных категориях, а именно – высокой степени абстракции, применении не алгоритмизуемых подходов при решении задач. Оба формата мышления заслуживают внимания и являются эффективными в своих областях знаний.

Но при столкновении данных форматов в процессе создания LegalTech-решений возникают противоречия ещё на первоначальных этапах реализации проекта. Традиционно любой подобный проект начинается с постановки цели, разработки концепции и подготовки ТЗ.

Если цель понимается разработчиком и юристом, как правило, одинаково, то уже при выработке концепции и методологии ее достижения зачастую начинаются существенные противоречия, не позволяющие перейти к написанию ТЗ.

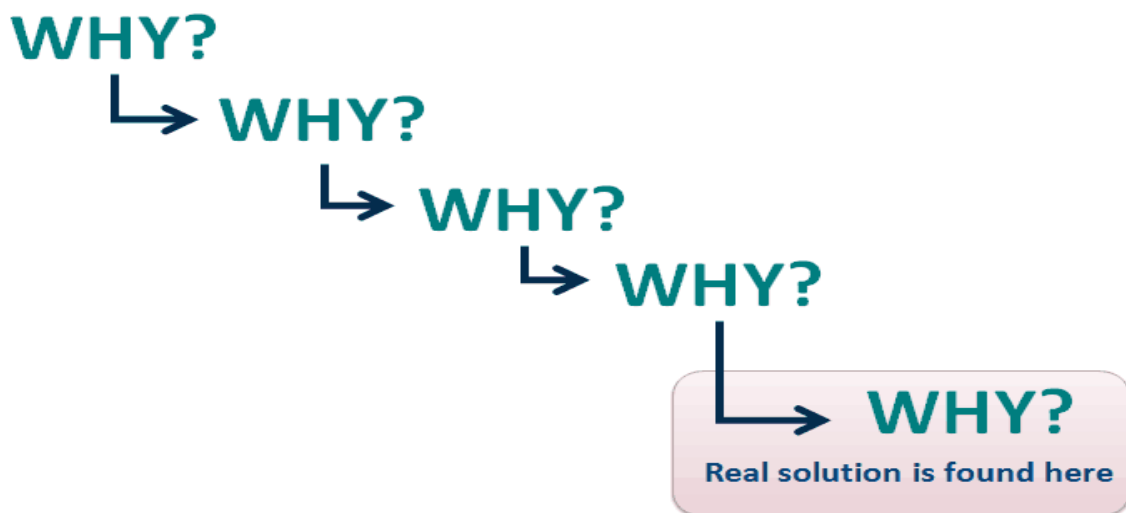
Когда мы говорим о создании продуктов для юридической сферы на основе инновационных технологий, для преодоления таких противоречий необходимо находить баланс и выстраивать эффективную коммуникацию между

представителями данных профессий. Для этого необходимо, чтобы и разработчики, и юристы не ограничивались собственными областями знаний, а стремились к пониманию особенностей деятельности друг друга и к поиску оптимальных решений существующих проблем для достижения общей цели.

Юристы должны стремиться к пониманию логики разработчиков и принципов работы существующих технологий, а разработчики – к пониманию базовых категорий юриспруденции, её основ и принципов.

При этом, на наш взгляд, большее влияние на сегодняшний день должны оказывать именно юристы, являющиеся носителями экспертных знаний предметной области разрабатываемых продуктов.

Эффективная коммуникация и взаимопонимание будут способствовать осознанию глубины существующих проблем в области LegalTech, которые могут быть решены при помощи метода first principles.



Такой подход – один из лучших способов деконструкции сложных проблем и раскрытия нестандартных возможностей. Его смысл заключается в том, чтобы разбить сложные проблемы на базовые элементы и затем реконструировать их снизу вверх.

Это один из лучших способов продвинуться от предсказуемых исходных данных к нелинейным результатам. Такой метод использовался философом Аристотелем, а теперь используется Илоном Маском и Чарльзом Мангером. Он позволяет преодолеть ложные убеждения и неэффективные аналогии, увидев возможности, которые все упускают. «В каждом систематическом изыскании (греч. Methodos), где существуют первые принципы, или причины, или элементы, знание и наука являются результатом познания этих принципов; мы считаем, что познали что-то, только узнав о первичных причинах, первичных первых принципах, вплоть до элементов», – писал Аристотель.

Данный метод отлично подходит и для решения задачи в сфере Legal AI, а именно – необходимо планомерно и пошагово реализовать продукт, продвигаясь от решения простых вопросов к сложным путем тесного взаимодействия между всеми участниками процесса разработки: юристами, разработчиками и лингвистами.

## 2.2. Структура знаний и восприятие данных

Ещё одним аргументом в подтверждение нашего подхода служат различия в восприятии данных специалистом соответствующей предметной области знаний и тем, кто с особенностями данной сферы незнаком.

Применительно к юриспруденции данные различия проявляются весьма наглядно.

Например, услышав термин «компания», не погруженный в юриспруденцию человек, вероятно, представит некоторую организацию, которая продаёт товар, выполняет работы или оказывает услуги.

Юрист, услышав данный термин, подсознательно выстроит для себя следующую картину:

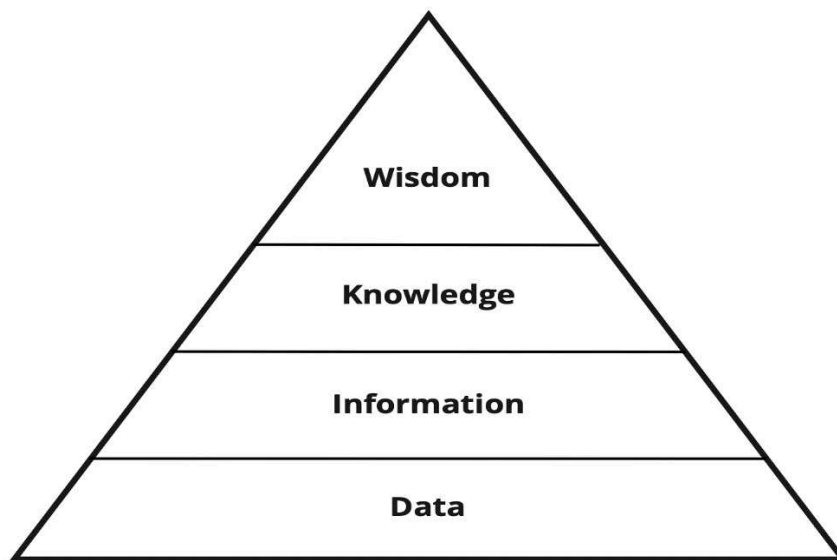
- есть некоторое юридическое лицо, которое функционирует в определённой области, является коммерческим образованием корпоративного типа;
- такое юридическое лицо равно имеет акционеров (участников), органы управления (как минимум общее собрание и единоличный исполнительный орган), которые осуществляют собственные полномочия в соответствии с законодательством и уставом;
- помимо этого, компания имеет фирменное наименование, юридический адрес, уставный капитал и др.

Данный список можно продолжать достаточно долго, но и приведённых примеров достаточно для понимания различий в глубине восприятия окружающего мира специалистами различных областей знаний.

Применительно к сфере LegalTech, в которой в рамках общего проекта по созданию юридических программных продуктов взаимодействуют юристы и разработчики, различия в понимании предметной области проявляются не только в глубине восприятия, но и во взглядах на одну и ту же проблему. Юристы стремятся понять сложность задачи с точки зрения соответствующей области знаний, а разработчики стремятся понять, каким образом можно описать программный продукт, решающий данную задачу.

Такие различия имеют в том числе научное обоснование.

Традиционно в теориях системного анализа информации существующие данные разделяют на иерархию уровней, каждый из которых своими свойствами существенно отличается от предыдущего. Наиболее распространенной моделью является DIKW (англ. data, information, knowledge, wisdom – данные, информация, знания, мудрость), предложенная в 1989 г. Расселом Акоффом (так называемая пирамида Акоффа).



В рамках данной иерархии выделяются:

1. data (неструктурированные, разрозненные данные);
2. information (структурированные данные, информация);
3. knowledge (знания);
4. wisdom (мудрость).

Перемещение между данными уровнями представляет собой процесс накопления знаний — обрабатывая больший объем данных и информации, субъект получает знания и понимание в рамках соответствующей предметной области. Чем больше знаний, тем на более высокой ступени пирамиды находится субъект.

Приведём простой пример перемещения между иерархией знаний внутри профессионального сообщества юристов:

- стажёр и помощник юриста, как правило, работают с информацией на уровнях data и information, обеспечивая её поиск и первичную обработку;

- юрист работает на уровне knowledge, обладая устойчивыми и системными знаниями предметной области, сформированным на основе длительной работы с информацией;
- старший юрист и выше работают на уровне wisdom, обладая накопленным опытом, мудростью и способностью видеть закономерности и взаимосвязи в мельчайших деталях.

Пирамида может быть представлена и в обратной последовательности – от мудрости к неструктурированным данным. Тогда становится очевидной иная закономерность – чем больше опыта и профессиональной мудрости у субъекта, тем меньше данных и информации ему необходимо для анализа ситуации и предоставления вывода.

Например, при разработке стратегии защиты интересов клиента в ходе судебного разбирательства старший юрист, как правило, безошибочно определит процессуальную категорию спора и проработает вопросы, которые входят в предмет доказывания по данному делу с учётом процессуального законодательства и сложившейся практики рассмотрения судом аналогичных разбирательств, соберёт необходимую доказательную базу, что в совокупности позволит эффективно достичь судебного решения в свою пользу. Менее опытный и квалифицированный специалист, вероятно, либо не выполнит необходимые мероприятия в полном объёме, либо выполнит лишние действия, не влияющие на решение судьи и конечный результат.

Другая не менее важная закономерность взаимосвязанных уровней заключается в том, что не специалист предметной области не сможет подняться выше уровня data (неструктурированные данные). Это вызвано в первую очередь тем, что для классификации данных и выделения из всего массива отдельной, значимой для дела информации требуются специальные познания.

Мы уже отмечали ранее, что одна из значимых компетенций юриста – это умение видеть в письменных документах именно существенные юридические факты для поиска возможных решений.

Решить данную задачу лицу, далёкому от юриспруденции, с высокой долей вероятности не удастся. Во многом именно в этом и заключается роль юриста. Данные тезисы в том числе ставят под сомнение практическую эффективность существующих на отечественном рынке платформ-агрегаторов общедоступной правовой информации (Спарк, Контр.Фокус и др.), которые позиционируются в качестве инновационных LegalTech-решений.

Как мы видим, данные продукты по-прежнему работают только на уровне неструктурированных данных, предоставляя пользователю всю найденную информацию, прямо или косвенно связанную с запросом, и не осуществляют её

обработку под призмой юридических знаний, оставляя задачу интерпретации этих данных человеку.

Приведём в качестве примера историю, основанную на реальных событиях. В нашей практике состоялся диалог с одной уважаемой компанией-разработчиком, представители которой занимаются разработками в сфере LegalTech. В ходе беседы представители компании поделились с нами планами по автоматизации проверки полномочий подписанта сделки.

Отметим, что мы также с удивлением для себя обнаружили, что в отделе, который занимается разработками LegalTech, нет юристов, они привлекаются «по необходимости». Логика представителей компании-разработчика была следующей:

- проверка полномочий не является сложной задачей, чтобы понять её специфику достаточно однократно обсудить вопросы с юристом (составить алгоритм, список вопросов и др.);
- суть проверки полномочий руководителя компании сводится по большей части к верификации его ФИО с данными из ЕГРЮЛ.

Однако любой юрист с данной логикой не согласится. На основании записи в ЕГРЮЛ мы не можем с полной уверенностью говорить о безусловном праве лица действовать от имени компании во всех случаях. Проверка полномочий руководителя хоть и не является трудной задачей, но требует, как минимум, комплексной проверки юридической связи между организацией и ее руководителем, а именно:

- проверки факта избрания руководителя высшим органом управления, срока полномочий и действительности такого решения;
- проверка факта вступления в должность руководителя юридического лица на основании приказа, совпадение всех идентификационных данных (ФИО, паспортные данные и др.) в каждом из документов;
- проверка полномочий руководителя на совершение юридически значимых действий (заключение сделок, выдача доверенности и др.) от имени организации на предмет наличия ограничений в уставе (например, для крупных сделок, сделок с заинтересованностью) и др.

Можем ли мы на примере данного сравнения говорить о корректности предложенного алгоритма проверки полномочий руководителя – вопрос риторический. К сожалению, в IT-сообществе распространён подход, когда реализация продукта начинается без привлечения экспертов из предметной области, либо с их привлечением на этапах тестирования.

Этой традиции уже не один десяток лет, и примеры мы можем найти даже в прошлом столетии. Так, в 1986 г. был проведён эксперимент по разбору британского

закона о гражданстве (The British Nationality Act) с применением логики языка Prolog. Суть эксперимента заключалась в попытке разложить текст данного закона на структурные единицы (смысловые триплеты), что и было в итоге сделано. Работы проводились без привлечения юристов, но в итоге эксперимент был признан успешным, несмотря на множественные недостатки и ограничения данной системы.

*«Our representation of the British Nationality Act was undertaken with no legal expert assistance. Our model of the Act expresses the layman's reading of the provisions. This in itself renders our British Nationality Act program of limited practical value. We could not use it in its present form for solving problems of British citizenship in actual legal practice».*

Вместе с тем практической ценности для юриспруденции данный эксперимент не принёс, что было признано самими разработчиками, поскольку без участия специалистов в области права невозможно воспроизвести смысловую структуру нормативного акта. Это во многом обусловлено особенностями юридического языка и терминологии, которая очень часто включает в себя неочевидные с позиции обыденного толкования значения.

В качестве примера можно привести следующий фрагмент из оригинальной статьи:

*«A complication that we anticipated was the presence of vagueness. The act contains such vague phrases as “being a good character,” “having reasonable excuse,” and “having sufficient knowledge of English.”»*

В данной ситуации авторы с позиции бытовой логики утверждают, что ряд терминов имеют размытое значение, основываясь, в первую очередь, на том, что нет какой-то одной статьи в законе, которая давала бы полную и точную дефиницию. Для практикующего юриста эта ситуация выглядит по-другому: рассматриваемые термины являются обширными, но не размытыми; по каждому из них накоплено несколько сотен лет судебной практики, где на многочисленных кейсах разбираются все возможные детали и нюансы.

В итоге эксперимент трансформации The British Nationality Act на язык Prolog был проведён ради самого процесса: была подтверждена возможность структурирования нормативных актов на языках формальной логики, но практическая значимость результата оказалась нулевой.

Таким образом, мы приходим к выводу, что при разработке решений в области LegalTech без ведущего участия профессиональных юристов обойтись невозможно. В противном случае нивелируется значимость и ценность таких решений для аудитории. Именно поэтому существующий сегодня фокус на работу с data (machine learning / deep learning) не позволяет продвинуться вперёд и решать задачи, в

которых требуется настоящая экспертиза на основе знаний и опыта конкретной предметной области.

В итоге мы пришли к выводу, что разработка Legal AI может быть выполнена только в результате создания независимого семантического блока, включающего в себя:

- структурированные модели юридических знаний (графы знаний и онтологии), воспроизводящие юридическую картину мира;
- набор методов и инструментов лингвистического анализа для процессинга текстов на естественном языке;
- элементы машинного обучения.

Данные инструменты в совокупности позволят создать качественно новые продукты в области LegalTech и перейти на следующие уровни работы в системе знаний (knowledge и wisdom).

### **2.2.1. Новый подход к формированию проектных команд**

Разработка решений в области Legal AI требует применения нового подхода к формированию бюджета проекта.

По нашей практике, что в структуре расходов основной статьёй (70-80%) является стоимость работы профессионалов предметной области (высоко специализированных юристов), обладающих глубокими знаниями и опытом на уровнях knowledge и wisdom, а также способных строить юридические онтологии, воспроизводя собственное экспертное видение юриспруденции. Остальные 20-30% расходов приходятся на оплату услуг разработчиков, обучение нейронных сетей, оборудование и др.

Обобщая изложенное, можно отметить, что наиболее дорогостоящими задачами являются строительство графа знаний предметной области и подготовка обучающего дата-сета, поскольку они требуют участия высокоспециализированных представителей предметной области.

Задачи по непосредственной разработке программного продукта менее затратны, однако с практической точки зрения разработка Legal AI не может быть осуществлена без решения обозначенных задач.

### 2.3. Неалгоритмизируемые процессы в LegalTech

Система знаний о какой-либо отдельно взятой предметной области может быть представлена в виде пирамиды из четырёх уровней (data>information>knowledge>wisdom), которая отражает как статическую глубину профессиональных познаний конкретного субъекта, так и динамический процесс накопления знаний.

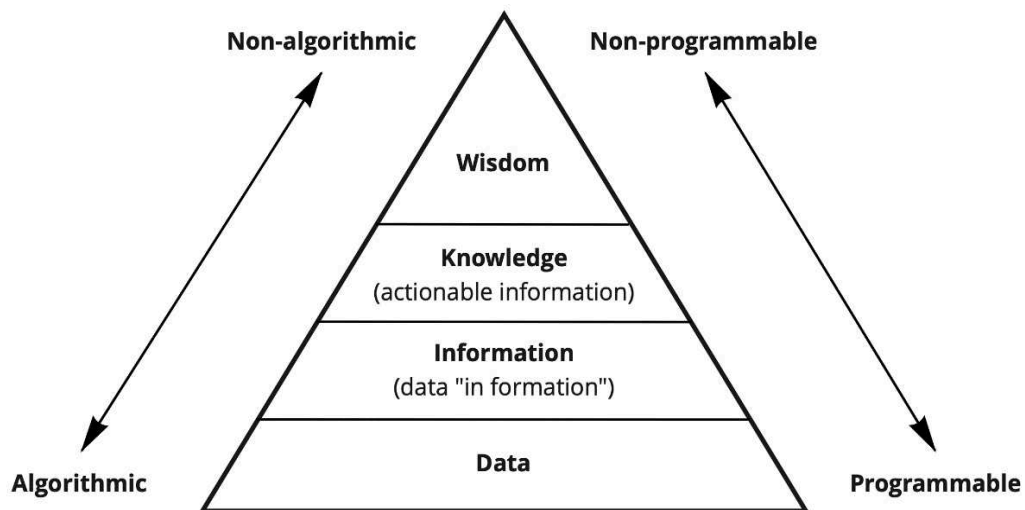
Основой и базисом всех знаний служат неструктурированные данные (data), из числа которых может быть выделена существенная информация (information). По мере накопления информации и ее обработки субъект формирует устойчивые знания предметной области, которые в последующем под призмой накопленного опыта трансформируются в мудрость (wisdom) и позволяют эффективнее работать с данными и информацией (knowledge).

Для достижения устойчивых знаний и мудрости требуется обширный опыт и глубокие экспертные познания в соответствующей области, поэтому данные уровни остаются недостижимыми для людей, не являющихся экспертами конкретной сферы. Применительно к сфере LegalTech это объясняет невозможность самостоятельного понимания IT-специалистами всех особенностей и деталей юриспруденции, поскольку они не владеют системными знаниями права и опытом их практического применения.

В юриспруденции (как в любой системе научных знаний) многие элементы взаимосвязаны и имеют между собой причинно-следственную связь.

Однако данная предметная область имеет и собственные особенности, благодаря которым такая связь может быть прямой (очевидной), косвенной (неочевидной) или отсутствовать вовсе. В связи с этим с точки зрения возможности алгоритмизации существует множество сложных областей и процессов, которые не поддаются программированию с помощью традиционных методов, применяемых в различных сферах.

Подобные особенности существуют не только в юриспруденции.



Наиболее наглядным примером служат эксперименты инженеров по созданию системы автопилота для легковых автомобилей, которая будет пригодна для автоматизированного управления транспортным средством в городской среде без участия человека.

Сегодня во всем мире в данном направлении достигнуты большие успехи, однако на первоначальных этапах разработчики столкнулись со следующей проблемой.

Изначальным подходом по алгоритмизации порядка управления автомобилем служила идеальная модель вождения, сформированная на основе всех правил дорожного движения.

В результате система управляла автомобилем только в строгом соответствии с правилами. На этапах тестирования такая модель показала свои существенные недостатки, которые связаны с тем, что помимо правил дорожного движения существуют отдельные неформализованные правила и законы вождения, а также общепринятые алгоритмы поведения водителей в нестандартных ситуациях, оцифровать которые было невозможно.

Например, водитель, приближаясь к пешеходному переходу и видя стоящего у него человека, на основе системы невербальной коммуникации способен понять и с точностью предугадать дальнейшие действия пешехода, а именно – намерен ли он переходить улицу или просто стоит у перехода без желания продолжать движение.

Система автопилота оказалась неспособна определять действия пешехода, стоящего возле нерегулируемого перехода, и, подчиняясь правилам дорожного движения, останавливалась и ждала пока пешеход не перейдет улицу.

Если пешеход не намерен переходить улицу, автопилот не поедет дальше, пока человек не уйдёт.

Другой пример – сложный нерегулируемый перекрёсток.

В стандартной ситуации водители пользуются жестами, с помощью которых могут разъехаться, не создавая заторы и аварийные ситуации.

Для системы автопилота, действующей на основе правил дорожного движения, такие невербальные методы были недоступны.

И, наконец, третий пример – слепая зона, расположенная после поворота, которая не видна водителю перед манёвром. Человек при управлении автомобилем в таких ситуациях действует на основе интуиции и когнитивного восприятия, предугадывая риск наличия или отсутствия пешехода на дороге после поворота, но у системы автопилота интуиции нет, поэтому такая ситуация на практике несёт в себе определённые риски.

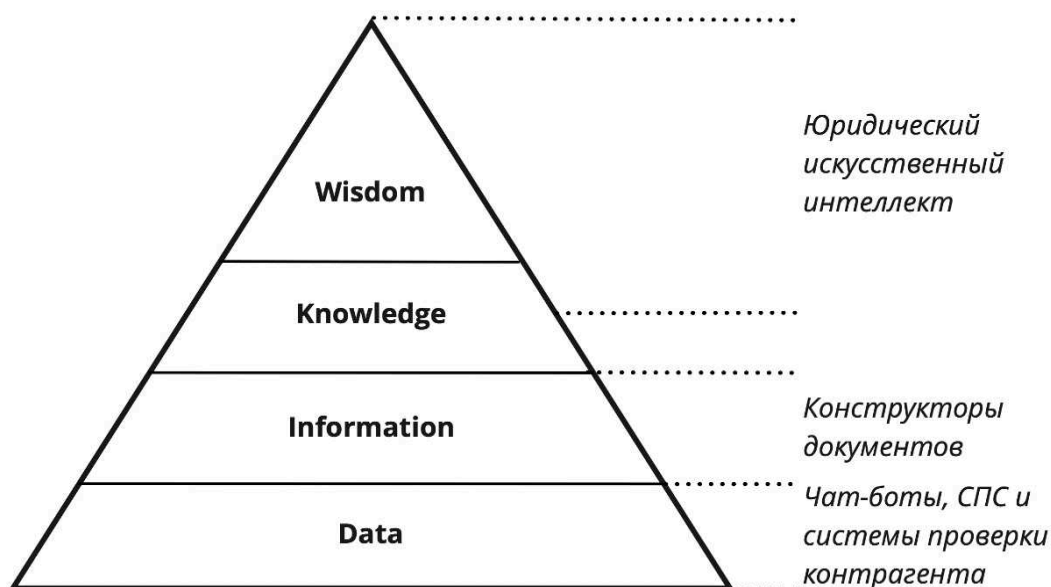
Описанные примеры и существующие неформальные правила и законы вождения, включающие в себя невербальную коммуникацию водителей, интуитивные действия и др. служат примером не алгоритмизируемой области, которая вырабатывается на основе длительной практики по управлению транспортным средством и с точки зрения пирамиды знаний находится на верхних уровнях (knowledge и wisdom).

Правила дорожного движения, в свою очередь, являются примером алгоритмизируемой области, которая располагается на нижних уровнях пирамиды (data и information). Аналогичные области существуют и в сфере юриспруденции, поэтому принятый у многих LegalTech-разработчиков подход, что все процессы в данной сфере могут быть алгоритмизированы традиционными методами, является заблуждением.

Кроме того, формат пирамиды и представления знаний позволяет наглядно определить место Legal AI в системе программных инструментов автоматизации юридической функции. Критерием разграничения инструментов выступает уровень в системе знаний, на котором функционирует тот или иной продукт, а именно:

- юридические чат-боты, справочно-правовые системы (Консультант Плюс, Гарант, Кодекс и др.) и системы проверки контрагента (Спарк, Контур.Фокус и др.) функционируют на уровне неструктурированных данных, обеспечивая поиск и аккумуляцию необходимых сведений;
- конструкторы правовых документов (Freshdoc, Easylaw, Doczilla и др.) функционируют на уровне информации, обеспечивая систематизацию необходимых данных, генерацию документов по шаблону;

➤ Legal AI, функционирующий на уровнях wisdom и knowledge, обеспечивает экспертную аналитику данных с точки зрения знаний предметной области.



## 2.4. Agile в LegalTech

Другой немаловажной проблемой является активная популяризация в IT-среде принципов и подходов Agile, которые в области LegalTech, по нашему мнению, создают очень большие сложности. Гибкая методология разработки (англ. Agile software development) — обобщающий термин для целого ряда подходов и практик, основанных на ценностях Манифеста гибкой разработки программного обеспечения и 12 принципах, лежащих в его основе, а именно:

- удовлетворение клиента за счёт ранней и бесперебойной поставки программного обеспечения;
- приветствие изменений требований даже в конце разработки (это может повысить конкурентоспособность полученного продукта);
- частая поставка рабочего программного обеспечения (каждый месяц или неделю, или ещё чаще);
- тесное и ежедневное общение заказчика с разработчиками на протяжении всего проекта;
- проектом занимаются мотивированные личности, которые обеспечены нужными условиями работы, поддержкой и доверием;

- рекомендуемый метод передачи информации – личный разговор (лицом к лицу);
- работающее программное обеспечение – лучший измеритель прогресса;
- спонсоры, разработчики и пользователи должны иметь возможность поддерживать постоянный темп на неопределённый срок;
- постоянное внимание улучшению технического мастерства и удобному дизайну;
- простота – искусство не делать лишней работы;
- лучшие технические требования, дизайн и архитектура получаются у самоорганизованной команды;
- постоянная адаптация к изменяющимся обстоятельствам: команда должна систематически анализировать возможные способы улучшения эффективности и соответственно корректировать стиль своей работы.

Применяется Agile как эффективная практика организации труда небольших групп (которые делают однородную творческую работу) в объединении с управлением ими комбинированным (либеральным и демократическим) методом.

Суть Agile заключается в том, что применяется итеративный подход при работе над проектом. Команда сконцентрирована на решении отдельных малых пользовательских кейсов в течение всего процесса, требования к конечному продукту формируются в динамике (по сути в процессе разработки).

Такой подход в итоге приводит к тому, что на первых этапах создаётся не сильно функциональный продукт, к которому в последующей перспективе интегрируются новые возможности. Данный подход имеет определённые преимущества, но в то же время имеет и существенные недостатки. К числу преимуществ традиционно относят увеличение скорости реализации проекта при сокращении ресурсов.

Критика Agile сводится к тому, что при нем часто пренебрегают созданием плана («дорожной карты») развития продукта, равно как и управлением требованиями, в процессе которого и формируется такая «карта».

Гибкий подход к управлению требованиями не подразумевает далеко идущих планов (по сути, управления требованиями в данной методологии просто не существует), а подразумевает возможность заказчика вдруг и неожиданно в конце каждой итерации выставлять новые требования, часто противоречащие архитектуре уже созданного и поставляемого продукта.

Такое иногда приводит к катастрофическим проблемам с массовым рефакторингом и переделками практически на каждой очередной итерации. Кроме

того, считается, что работа в Agile мотивирует разработчиков решать все поступившие задачи простейшим и быстреешим возможным способом, при этом зачастую не обращая внимания на правильность решения с точки зрения требований (подход – «работает, и ладно», при этом не учитывается, что может перестать работать при малейшем изменении или же дать тяжёлые к воспроизводству дефекты после реального внедрения). Это приводит к снижению качества продукта и накоплению дефектов.

Принципы Agile применимы при создании продуктов, ориентированных на массовую аудиторию, которые не предполагают глубокое внедрение в какую-либо предметную область, например, различные развлекательные сервисы, социальные сети и др.



В случае с LegalTech эта шутка имеет большую долю истины.

Как мы говорили ранее, успеха можно добиться только при неуклонном следовании *first principles*, когда ещё до перехода к работе программистов мы от и до понимаем, какие задачи необходимо решить, как их решить, и какой результат должен быть в конце.

Только при полном понимании всех деталей и требований к продукту мы можем перейти к его практической реализации. В противном случае велик риск упустить важные детали в алгоритме работы продукта, что в конечном итоге приводит к его невостребованности и неприменимости (*пациент не проснётся*).

### 3. Графовые системы и онтологии

#### 3.1. Концептуальное понимание картины мира

В юридической сфере помимо описанных факторов большую роль играет глубины понимания отдельных смыслов, концептов.

Русский язык, особенно с учётом профессиональной лексики, таит в себе опасную иллюзию, которая заключается в следующем.

Мы все говорим на одном языке, и кажется, что мы говорим об одном и том же, но на самом деле мы укладываем наши концепты в слова.

В процессе сжатия концепта в слово происходит его критическая деградация, которая становится обратимой только внутри профессионального сообщества.

Иными словами, два юриста с лёгкостью поймут друг друга, но юрист и разработчик – нет. Данные идеи подтверждаются теорией И.А. Мельчука «Смысл – текст».

Теория «Смысл  $\Leftrightarrow$  Текст» (ТСТ, или теория лингвистических моделей «Смысл  $\Leftrightarrow$  Текст», как её называют полностью) создана И.А. Мельчуком в середине 1960-х г. при активном участии ряда других лингвистов – прежде всего А.К. Жолковского, а также Ю.Д. Апресяна.

По замыслу её создателей, ТСТ является универсальной концепцией, которая может быть применима к любому языку.

На практике основным объектом для неё служил русский язык, а в 1980-е годы и в последующем теория разрабатывалась применительно к данным английского и французского языков.

Теория «Смысл  $\Leftrightarrow$  Текст» представляет собой описание естественного языка, понимаемого как устройство («система правил»), обеспечивающее человеку переход

от смысла к тексту («говорение», или построение текста) и от текста к смыслу («понимание», или интерпретация текста).

При этом приоритет в исследовании языка отдаётся переходу от смысла к тексту: считается, что описание процесса интерпретации текста может быть получено на основе описания процесса построения текста. Теория постулирует многоуровневую модель языка, то есть такую, в которой построение текста на основе заданного смысла происходит не непосредственно, а с помощью серии переходов от одного уровня представления к другому.

Помимо двух «крайних» уровней — фонологического (уровня текста) и семантического (уровня смысла), выделяются поверхностно-морфологический, глубинно-морфологический, поверхностно-синтаксический и глубинно-синтаксический уровни. Каждый уровень характеризуется набором собственных единиц и правил представления, а также набором правил перехода от данного уровня представления к соседним. На каждом уровне мы имеем дело с особыми представлениями текста – например, глубинно-морфологическим, поверхностно-синтаксическим и т.п.

Основной идеей данной теории является то, что одну и ту же мысль можно выразить с использованием различных синтаксических и лексических конструкций. Поэтому переводить текст в смысл в буквальном толковании невозможно, поскольку при выражении смысла устным и (или) письменным языком автор закладывает в него скрытые, понятные ему и лицам с аналогичными знаниями и мышлением смыслы.

Применительно к юридическому языку это демонстрируется на примере, который мы приводили выше (о различном понимании термина «компания» юристом и лицом без соответствующих знаний).

Данная теория в полной мере подтверждает вывод о том, что юрист и программист, обсуждая проекты в области LegalTech, говорят на разных языках и очень часто не понимают друг друга, что в итоге порождает конфликты и проблемы. Однако взаимодействие этих специалистов является ключом к успеху при создании Legal AI, поэтому важно стремиться к плодотворному и взаимному сотрудничеству.

### **3.2. Юридические концепты**

По итогам рассмотрения предыдущих вопросов мы пришли к выводу о сложности юридической области знаний и различной глубине понимания окружающего мира юристами и специалистами иных предметных областей.

Данная проблема находит своё практическое выражение даже в повседневной работе юриста.



Многие граждане и представители бизнеса нередко считают, что привлекать профессиональных юристов для сопровождения обычных и повседневных процедур, будь то оформление поставки партии товара или подписание очередного типового договора, не имеет смысла. К тому же с развитием LegalTech общественности становится доступно все больше различных электронных сервисов проверки контрагентов (Спарк, Контр.Фокус), онлайн-конструкторов договоров (Freshdoc, типовые формы в справочно-правовых системах «Консультант», «Гарант» и др.), позиционирующихся в качестве инновационных продуктов, способных заменить юриста в несложных, на первый взгляд, операциях (подробнее о существующем рынке LegalTech в России и сущности предлагаемых продуктов мы говорили в предыдущей статье).

Однако это очень большое заблуждение, которое нередко приводит к фатальным последствиям.

Для иллюстрации рассмотрим, что представляет собой проверка полномочий руководителя в обыденном понимании и в понимании юриста.

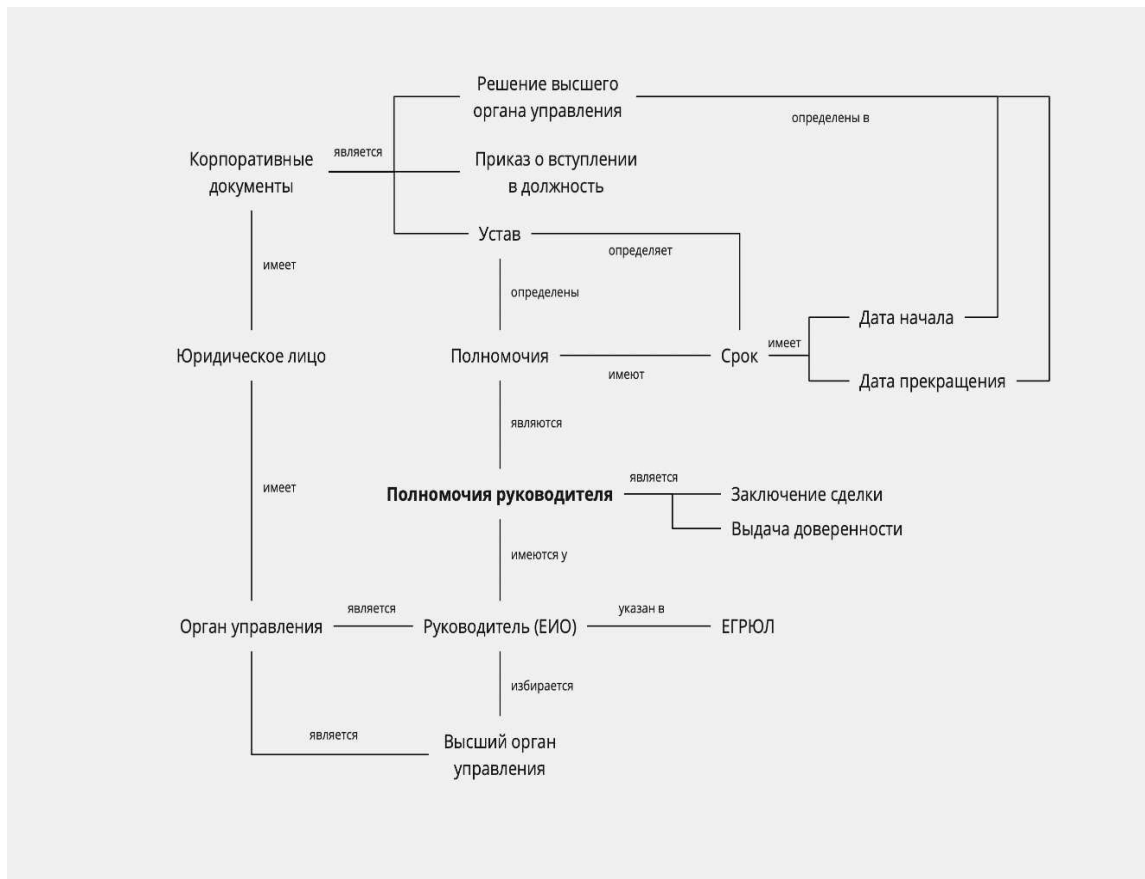
Предположим, что нам необходимо подписать с контрагентом договор о поставке товара. Контрагентом выступает организация, от имени которой действует генеральный директор. В подавляющем большинстве случаев лица без юридических

знаний сводят проверку полномочий к поиску контрагента в ЕГРЮЛ и верификации представителя компании со строкой «Лица, действующие от имени организации без доверенности». Если данные совпадают, значит договор можно подписывать.

Однако это не всегда так. С точки зрения юриста процедура проверки полномочий включает в себя анализ множества неочевидных и непонятных для обычного человека категорий, имеющих значение не только для подтверждения факта наличия полномочий у лица заключить сделку в данный момент времени, но и для обеспечения исполнения данной сделки и предупреждения оспаривания договора и признания его недействительным в последующем, что способно повлечь неблагоприятные последствия для всех сторон правоотношений.

При ответе на вопрос о наличии у лица (руководителя) полномочий подписать договор от имени организации юрист анализирует не только наличие статуса руководителя у лица на момент заключения сделки, но и проверку следующих факторов:

- право лица заключить сделку с конкретным объектом договора;
- наличие прав на соответствующий объект (право собственности на вещь, действительность права при заключении сделки с имущественными правами и др.);
- отсутствие признаков оспоримости сделки (непротиворечие сделки закону и иным нормативным актам, соответствие подписантов критериям полной дееспособности, вопросы аффилированности и взаимосвязанности, отсутствие признаков обмана, заблуждения, кабальности и др.).



Данный список может быть продолжен и далее, однако и приведённого выше достаточно для понимания, что проверка полномочий руководителя на заключение сделки существенно выходит за рамки поиска нужной фамилии в ЕГРЮЛ. Если ретранслировать данную ситуацию на тему Legal AI и автоматизацию проверки полномочий руководителя с помощью программных решений, мы сталкиваемся с необходимостью полного отражения данного концепта и фрагмента юридических знаний в данной области на языке формальной логики.

И для начала нам необходимо понять сущность алгоритма проверки полномочий, а именно: сколько шагов и какие действия необходимо выполнить, какие концепты и данные задействованы и др.

Только при условии решения данной задачи можно говорить о создании инструмента интеллектуальной проверки полномочий, которая, по существу, анализирует обстоятельства, не ограничиваясь обращением к ЕГРЮЛ. Если попытаться представить данный процесс в формате графа, он будет включать в себя сотни/тысячи узлов.

Другой пример – процесс подготовки проекта договора (например, договора поставки). Многие исходят из того, что для оформления большинства отношений, в том числе в предпринимательской сфере, не требуется участие юриста и достаточно использовать типовые шаблоны или конструкторы договоров, в которые

необходимо внести недостающие сведения: фамилии, наименования, реквизиты, название объекта и др., и договор можно отправлять на подписание.

## Шаблон. Договор поставки оборудования

## Опросный лист

 Нужна помощь в выборе настройки?

**Правовой статус сторон:**

Поставщик :

- Юридическое лицо;  
 Индивидуальный предприниматель;

Покупатель :

- Юридическое лицо;  
 Индивидуальный предприниматель;

**Принадлежности:**

- Перечень определяется в договоре;  
 Передаются в соответствии с законом.

**Документы:**

- Перечень определяется в договоре;  
 Передаются в соответствии с законом.

**Комплектность:**

- Определяется в договоре;  
 Определяется в соответствии с законом.

**Гарантийный срок использования Оборудования:**

- Определяется в договоре;  
 Качество Оборудования соответствует закону.

Договор содержит условие о монтажных и пусконаладочных работах:

- Да:  
 Стоимость работ входит в стоимость Оборудования;  
 Стоимость работ оплачивается Покупателем отдельно;  
 Нет.

Проводится обучение специалистов Покупателя по пользованию Оборудованием :

- Да:  
 Стоимость обучения входит в стоимость Оборудования;  
 Стоимость обучения оплачивается Покупателем отдельно;

## Список документов

500 Р



Разгрузите сотрудников.  
Избавьтесь от ошибок.  
Экономьте время и деньги.

**ДОГОВОР ПОСТАВКИ № № Договора**


Место заключения


Дата г.

Наименование Стороны, именуемое(ый, ая) в дальнейшем Поставщик, в лице Тип должности ФИО подписанта, действующего(ей) на основании Основание полномочий подписанта, с одной стороны, и

Наименование Стороны, именуемое(ый, ая) в дальнейшем Покупатель, в лице Тип должности ФИО подписанта, действующего(ей) на основании Основание полномочий подписанта, с другой стороны, вместе именуемые Стороны, а индивидуально – Сторона, заключили настоящий договор поставки (далее по тексту – Договор) о нижеследующем:

**1. Предмет договора**

1.1. В соответствии с Договором Поставщик обязуется передать Оборудование, наименование которого указано в п. 1.2 Договора (далее по тексту - Оборудование), в собственность Покупателю и провести монтаж Оборудования и пусконаладочные работы, а Покупатель обязуется принять и оплатить Оборудование в порядке и сроки, указанные в Договоре. 

1.2. В товаро-сопроводительном документе и Спецификации Оборудования (Приложение № № Приложения - Спецификация объекта договора к Договору), являющейся неотъемлемой частью Договора, Сторонами определены: 

наименование Оборудования  
 количество Оборудования  
 ассортимент Оборудования  
 принадлежности Оборудования  
 документы, передаваемые вместе с Оборудованием  
 комплектность Оборудования  
 стоимость Оборудования.

1.3. Поставщик гарантирует, что на момент заключения Договора, Оборудование принадлежит Поставщику на праве собственности, в споре и под арестом не состоит, не является предметом залога, в отношении Оборудования обременения правами третьих лиц отсутствуют.

1.4. Качество Оборудования, передаваемого по Договору, соответствует действующим в Российской Федерации стандартам и техническим условиям. Поставщик гарантирует Покупателю нормальную работу передаваемого Оборудования при условии соблюдения Покупателем инструкций по его технической эксплуатации.

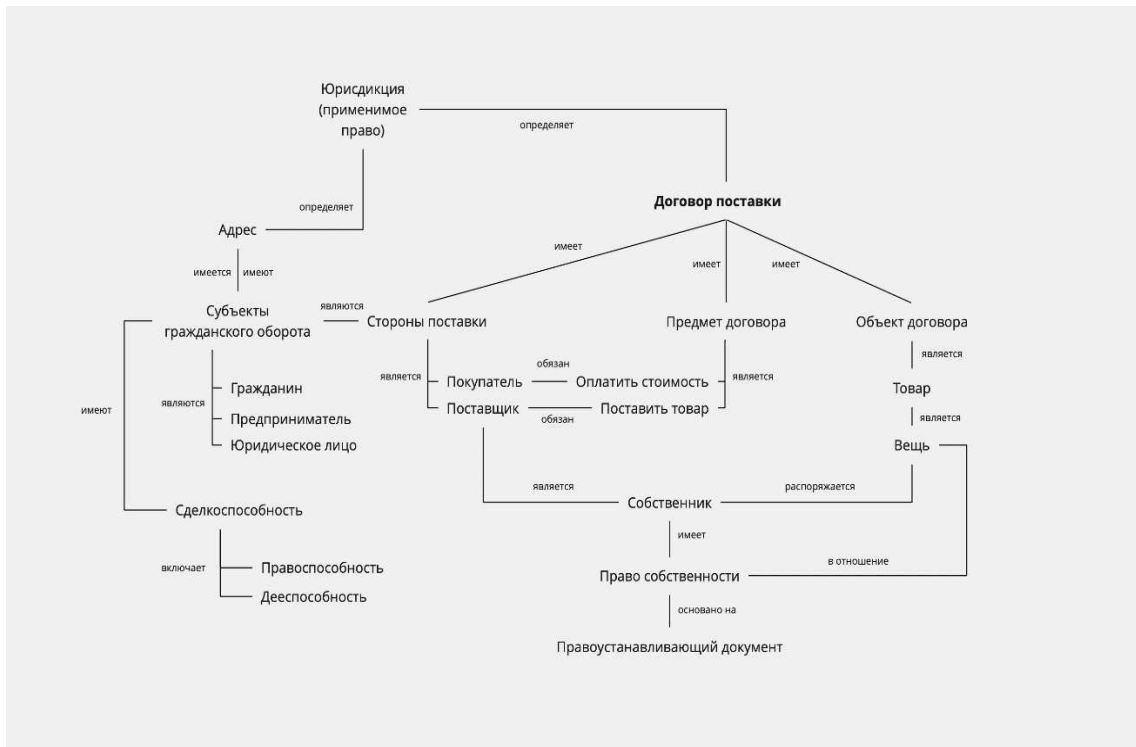
1.5. Стороны Договора определили, что техническая и коммерческая эксплуатация приобретаемого Оборудования должна обеспечивать его нормальное и безопасное использование в соответствии с целями покупки его по Договору.

1.6. Стороны согласовали необходимость упаковки Оборудования в тару, отвечающую требованиям ГОСТов, ТУ и обеспечивающую сохранность Оборудования при перевозке и хранении.

1.7. Гарантийный срок на Оборудование указан в Спецификации Оборудования.

Однако с точки зрения юриста данный процесс должен выглядеть совершенно иначе. Руководствуясь методом *first principles*, юрист должен ответить на следующие вопросы:

- в какой юрисдикции заключается договор и какому праву он подчинён;
- кем являются стороны договора с точки зрения их правового статуса (граждане, предприниматели, организации или публично-правовые образования, вопросы специальной правоспособности и др.);
- кем являются подписанты договора (представители по доверенности, законные представители, «органические» представители в виде единоличного исполнительного органа и др.);
- как подтверждается право собственности продавца на товар (наличие соответствующих документов и др.) и так далее.



Кроме того, юрист стремится индивидуализировать текст договора под обстоятельства клиента, формулируя условия, исходя из интересов и рисков стороны. Именно поэтому один и тот же договор поставки может быть подготовлен и на 5, и на 60 страниц. Безусловно, многие юристы пренебрегают этим и сознательно в ряде случаев не следуют описанному алгоритму.

Поэтому распространённая и в целом приемлемая практика – это, когда многие важные этапы опускаются в целях оптимизации всего процесса (для простых и незначительных сделок, где риски минимальны и ими можно пренебречь).

Другой случай – юристы ленятся и либо сознательно пропускают данные проверки, либо осуществляют их поверхностно и сугубо формально. ФНС России неоднократно разъясняла правила должной осмотрительности при выборе контрагента и критерии оценки его добросовестности.

Однако далеко не многие юристы руководствуются данными рекомендациями при сопровождении сделок и проверке полномочий её подписантов, что является неприемлемой, но, к сожалению, распространённой практикой.

Данные примеры наглядно иллюстрируют различное понимание стандартных в деловой практике процедур с точки зрения юридической логики и логики специалистов других областей знаний. Многие, сталкиваясь с необходимостью подготовить проект договора или проверить полномочия подписанта сделки, на подсознательном уровне полагают, что это предельно простые задачи, требующие не более 1-2 часов времени.

Однако это далеко не так. Осознание глубины проблемы и отказ от подобных убеждений – первый шаг на пути к созданию Legal AI.

Третий пример, иллюстрирующий глубину и концептуальность понимания юристами внешних обстоятельств (событий и действий как разновидностей юридических фактов), — это вопросы юридической ответственности. Юридическая ответственность представляет собой меры государственного принуждения, применяемые к лицу за совершение противоправного деяния.

Наиболее сложным видом ответственности является ответственность лиц, входящих в состав органов управления организации.

В зону риска попадают контролирующие лица, которые в силу своего служебного положения (осуществление полномочий члена совета директоров, единоличного исполнительного органа) в ходе хозяйственной деятельности организации принимают решения и заключают сделки, которые имеют определённую степень предпринимательского риска. В силу данного факта такие лица должны действовать добросовестно и разумно, поскольку от их решений зависит будущее финансовое положение компании.

За нарушение данных требований они могут быть привлечены как к имущественной ответственности за действия, которые повлекли за собой неплатёжеспособность или убытки компании, так и к дисциплинарной (при исполнении обязанностей на основе трудового договора), административной и уголовной ответственности при наличии в их действиях составов административного правонарушения или преступления.

Приведём пример наступления неблагоприятных последствий при недобросовестных действиях контролирующих лиц.

Генеральный директор регулярно заключал договоры поставки в пользу компании партий товаров, стоимость которых в 1,5 раза превышала рыночную стоимость аналогичных товаров у других поставщиков.

Увеличенные расходы на закупку товаров генеральный директор объяснял бенефициарам компании высокой надёжностью поставщика и наличием устойчивых и длительных партнёрских отношений с ним, за счёт чего условия поставки могли включать в себя и рассрочку оплаты, и доставку товаров на отдалённые производственные объекты и др.

Однако фактически генерального директора и учредителя поставщика связывали личные приятельские отношения, благодаря которым поставщик получал необоснованную прибыль за счёт превышения цены по сравнению с рыночной на протяжении нескольких лет. Полученная в результате сверхприбыль в последующем была выведена в форме дивидендов и неформально распределена поровну между генеральным директором покупателя и учредителем поставщика, а компания-покупатель на протяжении многих лет несла убытки в связи с переплатой за товар.

Как только данные факты стали известны участникам компании-покупателя, генеральный директор был привлечён к дисциплинарной ответственности в виде увольнения на основании п. 9 ст. 81 ТК РФ (принятие необоснованного решения руководителем организации, повлёкшего за собой нарушение сохранности имущества, неправомерное его использование или иной ущерб имуществу организации).

Однако убытки компании в последующей перспективе повлекли за собой существенное ухудшение платёжеспособности, что в результате привело к невозможности исполнения обязательств перед иными контрагентами и работниками предприятия.

В результате было инициировано дело о банкротстве, в рамках которого действия бывшего генерального директора привлекли внимание конкурсного управляющего.

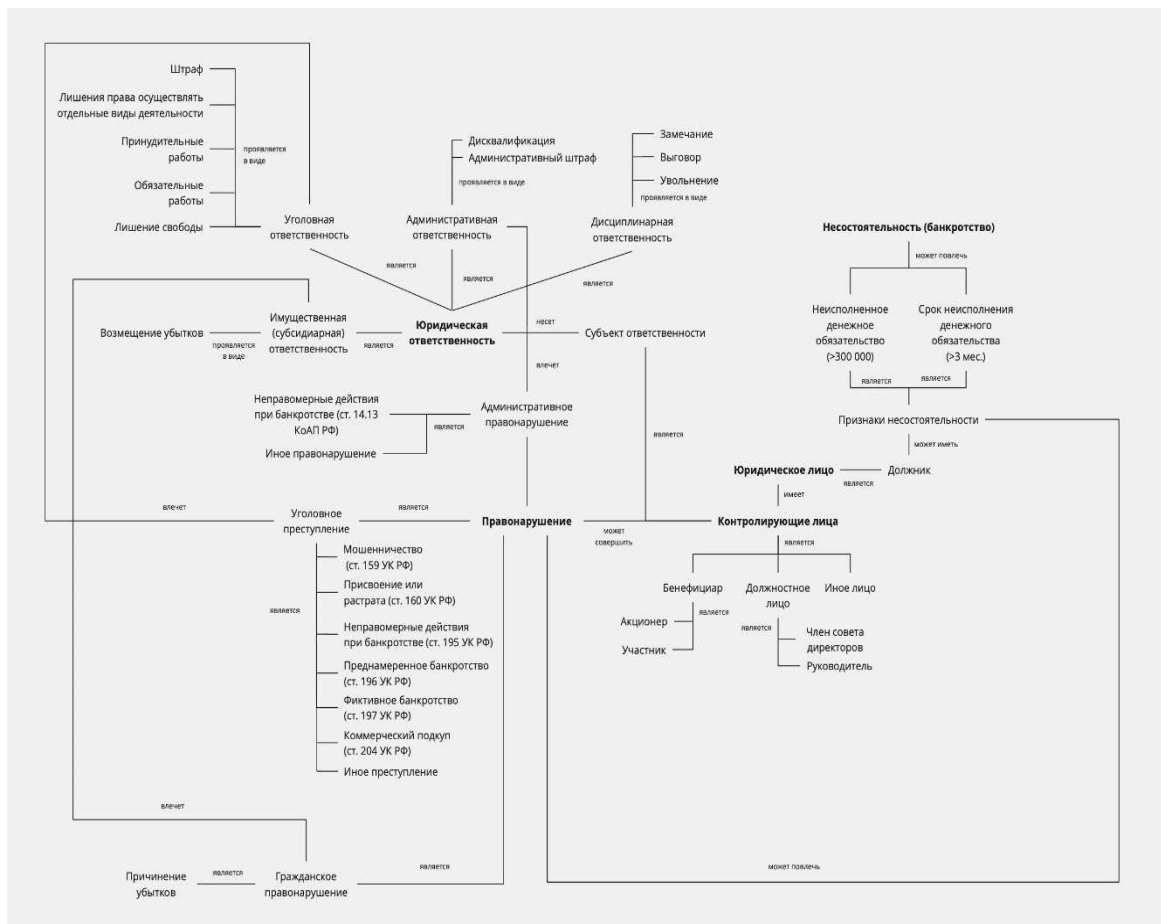
В отношении генерального директора было возбуждено уголовное дело, по итогам расследования которого виновные лица были привлечены к уголовной ответственности за совершение преступлений, предусмотренных ст. 160 (присвоение вверенного виновному имущества) и ст. 204 (коммерческий подкуп) УК РФ.

Помимо этого, в рамках уголовного дела к генеральному директору был предъявлен гражданский иск о возмещении убытков юридическому лицу в размере необоснованной переплаты по поставкам.

С точки зрения генерального директора данная схема выглядела соответствующей закону, поскольку в его понимании закупка товаров по завышенной цене у проверенного поставщика была платой за надёжность контрагента, а получаемое скрытое вознаграждение – благодарностью поставщика за долгосрочное сотрудничество.

Однако для юриста такие риски являются прогнозируемыми ещё задолго до их реализации. Понимая и оценивая риски, юрист подходит к анализу каждого принимаемого решения, будь то заключение сделки или проверка контрагента, с точки зрения возможных неблагоприятных последствий в виде мер юридической ответственности, поскольку понимает механизмы реализации таких рисков и специфику коммерческих взаимоотношений. Закупка товара по завышенной цене сама по себе влечет множество вопросов и несёт в себе значительное число рисков в виде ответственности для контролирующих лиц. То есть в юридической парадигме любое действие прямо или опосредованно влечет за собой определённые правовые последствия. Одно действие может стать основанием для привлечения к различным видам ответственности.

Благодаря пониманию таких взаимосвязей и закономерностей профессиональный юрист оценивает ситуации и осуществляет собственную экспертизу.

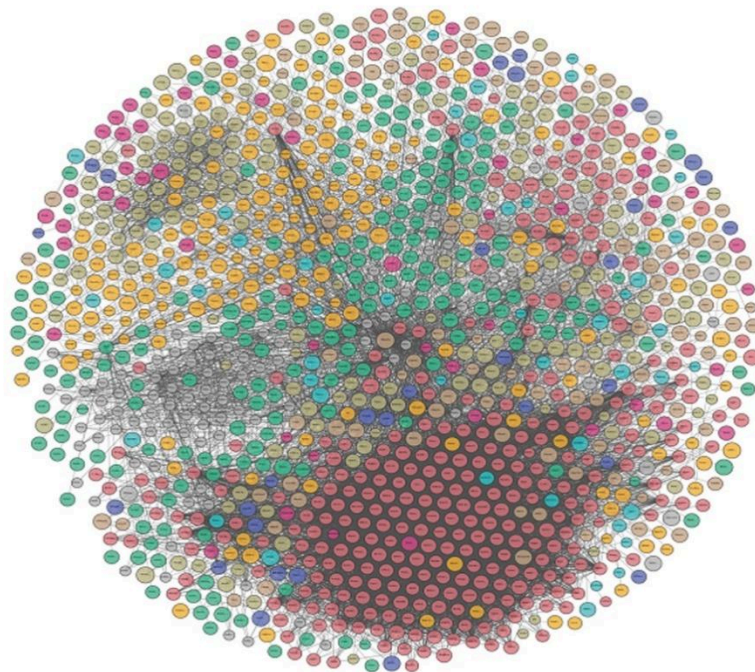


Именно такое понимание функциональных задач мы вкладываем в Legal AI. Настоящим технологическим прорывом в области LegalTech станет создание инструмента, позволяющего оценивать юридические факты с точки зрения профессиональной логики, видеть взаимосвязи и закономерности, очевидные для юриста, но скрытые для обывателя.

Это станет большим шагом в сторону автоматизации юридической функции, что при существующих продуктах на рынке LegalTech пока остаётся невозможным.

### 3.3. Онтологии и графы знаний

В процессе разработки программных решений для автоматизации юридической деятельности на основе искусственного интеллекта важным этапом, обеспечивающим возможность достижения успешного результата, является создание унифицированной базы знаний, отражающей особенности юридической картины мира.



Вопросы о необходимости структурирования и систематизации знаний в различных предметных областях не являются новыми. Ещё в начале 2000-х годов в

Европе начались разработки первых онтологий, а также были выработаны единые стандарты их построения.

В России единый подход не выработан, поэтому развитие онтологических систем знаний происходит менее активно.

Онтология является унифицированной и структурированной базой знаний необходимой предметной области, представляющей собой объективное семантическое отражение картины мира в структурированном формате. Онтология включает в себя набор связанных между собой терминов, иерархически записанных в формате классов, подклассов различного уровня и связей (Relationships) между ними, что позволяет соотносить имеющиеся данные между собой с точки зрения экспертной логики.

Среди способов систематизации знаний онтология является гораздо более продвинутой моделью, чем таксономия.

Таксономией является любая структура знаний в виде иерархически соотносящихся сущностей.

От онтологии её отличает наличие связей между структурными единицами только 1 типа – родительский или дочерний элемент («subclass of»). Онтология, в отличие от таксономии, обладает большим количеством связей между всеми структурными единицами, представляя собой способ формализации знаний, абстрактных или специфических, реализованный на основе формального описания объектов, фактов и отношений между ними.

Например, если в таксономии класс «Акционерное общество» может иметь только дочернюю связь с классом «Юридическое лицо» (акционерное общество является разновидностью юридического лица), то в онтологии «Акционерное общество» будет иметь сотни и тысячи связей, отражающих взаимосвязь акционерного общества с субъектами, объектами права и др. Данное ключевое отличие говорит о том, что таксономии не позволяют отвечать на юридические вопросы в силу отсутствия в их составе важных элементов в виде связей.

В формате онтологии крайне сложно выражать такие важные для юриспруденции конструкции как время (процессуальные сроки, сроки исполнения обязательств и др.) и модальность. В случае со временем мы вынуждены иметь дело с громоздкими конструкциями (например, как на изображении ниже), с которыми сложно работать на практике.

В случаях с модальностью на языке OWL к настоящему времени не решена проблема выражения таких состояний, как отрицание, сомнение, различного рода субъективные факторы и оценочные категории.

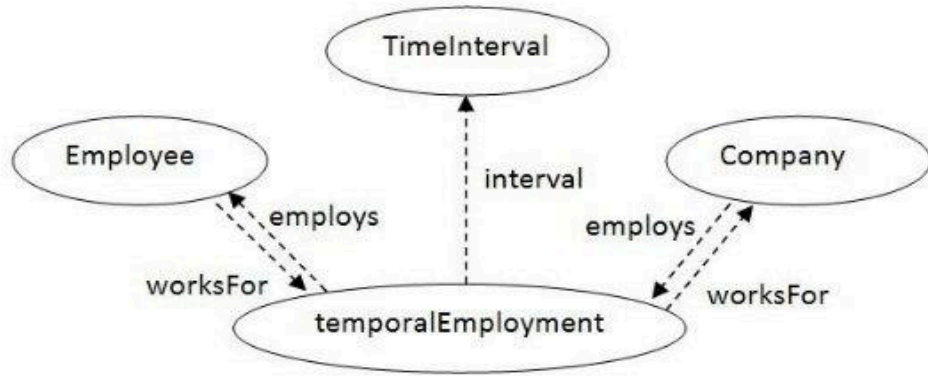


Fig. 2. Example of N-ary Relations

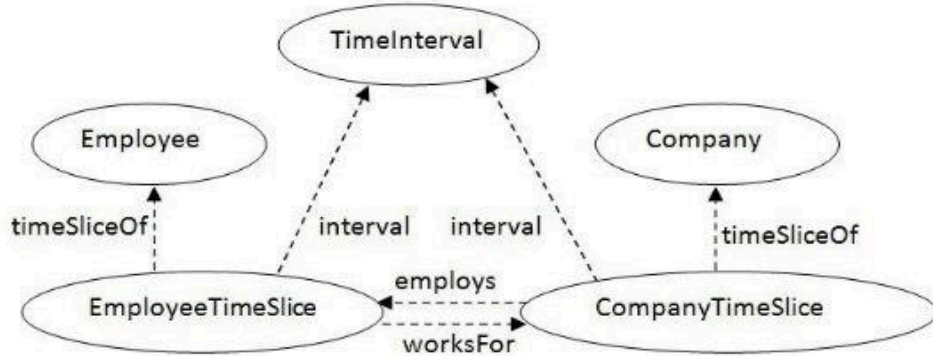


Fig. 3. Example of 4D fluents

Идеальной моделью представления знаний какой-либо предметной области является граф знаний.

Граф структурно включает в себя графовые хранилища семантических метаданных и онтологий, которые в данном случае выступают в роли полуструктурированной модели предметной области, являясь ядром графа знаний.

В результате такого способа обеспечивается возможность решения интеллектуальных задач с помощью постоянной циркуляции данных за счёт применения методов машинного обучения.

Говоря о соотношении графа знаний и онтологии, необходимо отметить, что онтология выступает в качестве способа формализации знаний в формате графа.

Как верно отмечал Д. Муромцев, для графов знаний онтология – это семантическая основа представления данных, базирующаяся на логике и включающая терминологический словарь и набор утверждений о моделируемых объектах.

В результате граф может иметь в себе в качестве семантической основы множество онтологий, обеспечивая комплексную концептуализацию всех знаний предметной области.

Онтологии и графы, создаваемые для Legal AI, должны иметь в совокупности сотни и тысячи классов и связей для решения даже простых юридических задач.

При этом решающее значение имеет именно количество связей, отражающих отношения между классами.

Для приблизительной оценки необходимого количества связей в нашей практике мы ориентируемся на полный граф (где каждая пара различных вершин смежные) и в результате получаем, что необходимое количество связей на порядок превышает количество необходимых классов:

Оптимальное количество связей  $\approx n*(n-1)/2$ , где  $n$  – количество задействованных классов, вершин.

Обращаясь к зарубежному опыту, нам удалось найти глубоко детализированные онтологии, применяемые в медицине (онтология геномов и др.), финансовой сфере и др. В области права существуют такие онтологии, как FOLaw, FBO, LKIF, Legal Rule ML и др.

В России таких примеров найти не удалось, не говоря даже конкретно о юридических онтологиях.

Исследование вопросов, связанных с онтологиями, обзор существующих зарубежных наработок в области юридических онтологий и их значение мы отдельно рассмотрим в следующей статье, но некоторые особенности, необходимые для понимания глубины проблемы, будут описаны далее.

Одним из наиболее впечатляющих примеров существующих онтологий является FIBO (Financial Industry Business Ontology), разрабатываемая в настоящее время международным сообществом под управлением консорциума OMG, который занимается объектно-ориентированными технологиями и стандартами.

Стандарты FIBO применяются многими налоговыми органами различных государств в том числе в рамках автоматического обмена информацией (сведения о бенефициарах и др.).

Онтология FIBO имеет своей целью воспроизвести структуру отношений между участниками финансового рынка по поводу различных финансовых инструментов.

Многими специалистами данной области FIBO рассматривается в качестве бизнес-онтологии, тем самым позиционируется её широкая и универсальная сфера применения.

Среди мнений встречаются также позиции о том, что бизнес-сфера включает в себя (поглощает) и сферу права, а также что унификация знаний в области финансов является основой для построения общих юридических концептов.

Мы считаем, что обе позиции являются заблуждением, в подтверждение чего далее данная онтология FIBO будет рассмотрена более подробно.

В онтологии FIBO содержится семантически связанные понятия финансовой отрасли с описанием их смысла и практики использования специалистами.

В частности, в FIBO описаны такие базовые понятия, как юридические лица, рыночные данные и финансовые процессы, структура и договорные обязательства, различные финансовых инструментов и др.

Классы и подклассы онтологии FIBO представлены двумя способами: формальным описанием понятий и их взаимосвязей на языке OWL, а также их описанием на естественном языке с использованием толковых словарей финансовой отрасли.

Предполагается, что онтология FIBO должна стать общим языком для финансовой индустрии, поддерживающим автоматизацию бизнес-процессов. Она предназначена для использования разработчиками, бизнес-аналитиками и другими участниками сферы финансов.

Бизнес-термины и определения, описанные в FIBO, могут быть использованы в качестве эталонной модели, с которой финансовые организации могут связывать свои собственные (локальные) модели.

Появляется возможность создавать логические модели данных, которые получают из FIBO свою формальную семантику.



В действительности FIBO является не одной онтологией, а набором большого количества онтологий, которые разделены по модулям и подмодулям.

Модули (подмодули) включают наборы совместно используемых онтологий. Между собой онтологии FIBO связаны отношением «использует» – одна онтология может использовать, расширять и уточнять понятия, описанные в другой онтологии (или наборе онтологий).

FIBO на очень глубоком уровне детализирует финансовые инструменты и термины, актуальные для финансовой отрасли.

Но финансовая отрасль во многом пересекается с юриспруденцией: финансовые сделки являются разновидностями частно-правовых договоров (обязательств), а финансовые инструменты – объектами гражданского оборота (ценные бумаги и др.).

Поэтому данная онтология содержит в том числе правовые категории, необходимые в той или иной мере для финансовой экспертизы, однако уровень их детализации является невысоким, что не позволяет применять FIBO для целей юриспруденции.

## ОБЩАЯ СТРУКТУРА FIBO

<b>Онтология верхнего уровня</b>		
<b>Базовый модуль FIBO (FIBO Foundations): верхний уровень абстракции</b>		
Модуль бизнес-сущностей ( <b>FIBO</b> , Business entities)	Модуль финансово-го бизнеса и коммерции ( <b>FIBO</b> , Financial business and commerce)	Модуль индексов и показателей ( <b>FIBO</b> , Indices and indicators)
<b>Онтологии договоров</b>		
Онтология ценных бумаг (основных, акций)	Онтология ценных бумаг (долговых)	Онтология производных финансовых инструментов
Онтология займов	Онтология капиталов	Онтология варрантов
Ценообразование и аналитика		
Процессы (выпуск ценных бумаг и акций)		
Будущее <b>FIBO</b> : понятия, связанные с отдельными учреждениями, их требованиями к отчетности		

Для понимания различного уровня детализации сферы финансов и права приведём несколько фрагментов онтологии FIBO.

- ▼ ○ document
  - certificate
  - ▼ ○ legal document
    - ▼ ○ contract document
      - ▼ ○ insurance policy
        - bond insurance
    - government-issued license
    - ▼ ○ identity document
      - birth certificate
      - driver's license
      - passport
    - ▼ ○ license
      - business license
      - government-issued license
    - municipal debt funds usage
    - municipal debt source of funds
    - ▼ ○ offering document
      - offering statement
      - private placement memorandum
      - prospectus
    - ▼ ○ transaction confirmation
      - over-the-counter (OTC) transaction confirmation
      - swap confirmation
  - notice
  - ▼ ○ publication
    - ▼ ○ catalog
      - financial product catalog
    - economic news release
    - ▼ ○ offering document
      - offering statement
      - private placement memorandum
      - prospectus
    - ▼ ○ published financial information
      - ▼ ○ market rate
        - daily average market rate
        - end-of-day market rate
      - ▼ ○ quoted exchange rate
        - currency forward rate
        - ▼ ○ currency spot rate
          - currency spot buy rate
          - currency spot mid rate
          - currency spot sell rate
      - ▼ ○ reference interest rate
        - base rate
        - ▼ ○ interbank rate
          - interbank bid rate
          - interbank mid rate
          - interbank offered rate
        - overnight rate
    - quoted price
  - reference document
  - ▼ ○ report
    - ▼ ○ assessment report
      - appraisal
      - rating report
    - regulatory report

На данном фрагменте отражены документы, используемые в сфере финансов (порядка 57 категорий).



На данном – известные FIBO источники права (порядка 6 категорий).

Вследствие различия глубины проработки вопросов финансов и права онтология FIBO малоприменима для решения задач в сфере Legal AI.

Кроме того, существуют и другие объективные причины, ограничивающие возможность применения онтологии FIBO для целей отечественного юридического искусственного интеллекта.

### 3.3.1. Успешный опыт построения индустриальных графов знаний

Ранее мы приводили в качестве примера успешной и глубоко детализированной онтологии разработку онтологии генов в рамках исследований в области молекулярной биологии.

Для её создания были вложены значительные финансовые ресурсы, и в результате научное сообщество получило функциональный инструмент для дальнейших исследований.

Данная онтология воспроизводит концепцию генома, включающую в себя как функциональное описание генов (функция и роль того или иного гена), так и их структурное описание (химический состав гена).

Такая структурированная модель знаний позволяет, во-первых, интегрировать знания из различных баз данных в едином формате, во-вторых, генерировать выводы о функциональности вновь открытых генов и получать представление о сохранении и дивергенции биологических подсистем. Онтология генов имеет свойство универсальности.

Существующие в молекулярной биологии элементы одинаковы и равным образом признаются учёными вне зависимости от территории, страны и др. Отличается лишь их индивидуальный набор у каждого представителя того или

иного вида организмов. Данные особенности предметной области позволяют практически применять подобную базу знаний во всем мире.

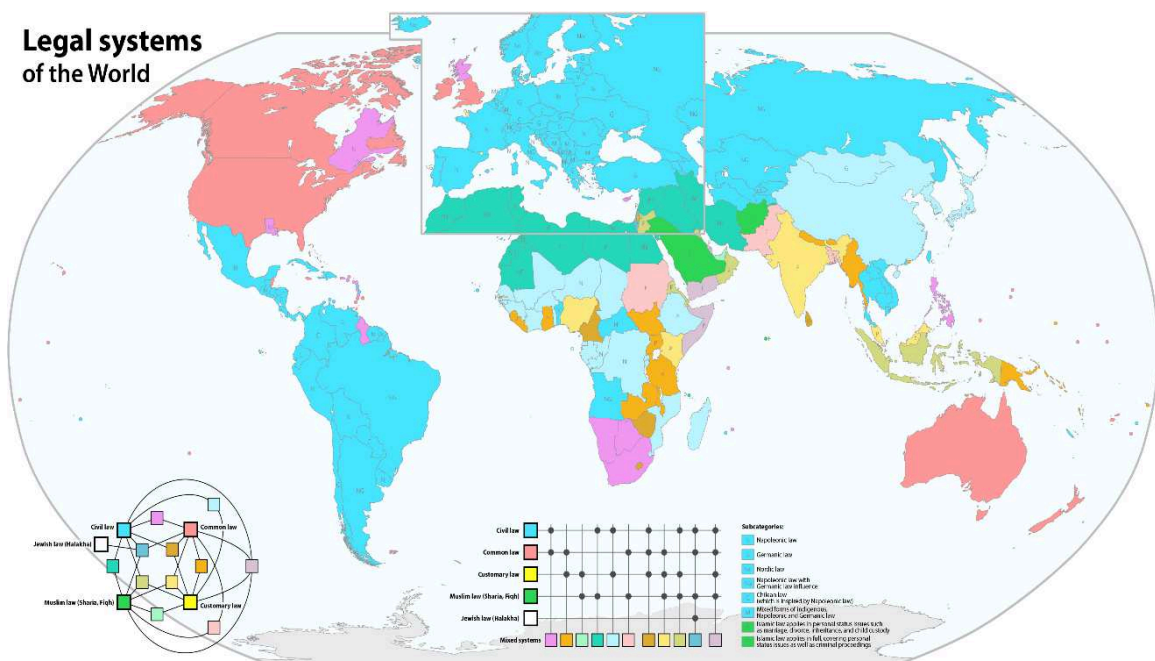
В области юриспруденции создать общеприменимую модель знаний невозможно.

Первая причина – фундаментальное различие правовых систем.

В теории права принято выделять 5 правовых систем: континентальную (например, Германия, Франция, Россия и др.), англо-американскую (США, Великобритания, Канада, Австралия), социалистическую (Китай, КНДР, Куба и др.), а также религиозную (основана на имплементации в право религиозных канонов и их законодательное развитие) и традиционную (основана на обычаях и общинных правилах поведения).

Каждая из правовых систем имеет фундаментальные отличия в принципах построения права и законодательном регулировании общественных отношений.

Крупнейшими в мире по числу представителей являются континентальная (романо-германская система) и англо-американская (система общего права) правовые семьи.



Однако даже внутри одной правовой системы существуют множественные ветви, характеризующиеся наличием индивидуальных особенностей, поскольку право в каждом государстве формируется под воздействием множества факторов (экономических, исторических, социальных, географических и др.).

Например, Россия и Германия относятся к одной правовой семье, однако их законодательство разительно отличается, хотя и существуют отдельные схожие институты.

Так, в гражданском праве и России, и ФРГ в качестве отдельного вида договора о передаче имущества в собственность является договор купли-продажи недвижимости как соглашение продавца и покупателя, по которому продавец обязуется передать в собственность покупателя недвижимую вещь, а покупатель – принять и оплатить ее стоимость.

Однако есть существенное отличие в понимании сущности данных отношений и в правовых последствиях их реализации.

В России договор купли-продажи недвижимости является обязательственной сделкой, которая создаёт взаимные обязательства между двумя сторонами.

Право собственности на недвижимость переходит к покупателю после государственной регистрации такого перехода.

Если договор купли-продажи признается недействительным, то в качестве последствий применяется двусторонняя реституция (взаимный возврат всего полученного каждой из сторон).

В ФРГ договор купли-продажи включает в себя две сделки: обязательственную и вещную. По обязательственной сделке создаются взаимные обязательства, а по вещной – переходит право собственности на вещь.

В результате право собственности переходит к покупателю с момента заключения вещной (распорядительной) сделки. При этом в силу принципа абстракции если обязательственная сделка признается недействительной, то это не влияет на действительность сделки вещной, и право собственности за покупателем сохраняется.

Вторая причина – фундаментальные различия в языках.

Всего в мире насчитывается более 7 тыс. языков, из которых только 40 являются самыми распространёнными для 2/3 населения Земли.

Безусловно, английский язык является наиболее распространённым в том числе как государственный язык, на котором составляются источники права в различных странах. Практически любой текст можно перевести с одного языка на другой.

Однако не любой текст в результате такого перевода сохранит 100% своего смыслового содержания и будет тождественным. Данная особенность крайне актуальна и для юридических текстов, которые имеют собственную специфику в виде терминологии, уникальных значений и др.

Приведённые причины позволяют прийти к выводу о невозможности создания общеприменимой структурированной базы знаний в области права и необходимости её самостоятельной разработки в отдельной стране с учётом её государственного языка и права.

Право государств отличается настолько фундаментально (даже внутри одной правовой семьи), что применение зарубежных разработок для решения задач LegalTech в России становится невозможным. Равным образом невозможно использование существующей онтологии FIBO, поскольку она создана на английском языке представителями англо-американской правовой семьи. Это требует выработки самостоятельного подхода и создания уникальной онтологии, воспроизводящей право РФ.

### **3.3.2. Количество классов и связей в практических онтологиях**

Онтология представляет собой структурированную модель представления системы знаний какой-либо предметной области на основе описания объектов (классов), их свойств и взаимосвязей с другими объектами. Построение такой модели позволяет систематизировать знания в едином формате и использовать их в том числе для генерации выводов. Однако функциональное назначение онтологий существенно шире.

В мировой практике существует множество примеров онтологий, которые создаются для различных целей (например, упомянутые нами FIBO в финансовой области, онтология геномов в молекулярной биологии и др.).

Существующие на текущий момент онтологии преимущественно создаются и применяются для обмена данными между различными субъектами, используя такую модель представления данных в качестве метаязыка, упрощающего и ускоряющего взаимообмен информацией.

Однако для обеспечения такого функционала онтологии создаются с упором на универсальность и возможность многократного применения (свойство «reusable»). Это влечёт за собой необходимость построения онтологий с высокой степенью абстракции, которая проявляется в максимальном обобщении и сокращении количества классов.

В результате такие онтологии обеспечивают возможность обмена данными, но становятся непригодными для использования в качестве базы знаний в виду неглубокой детализации.

Другая проблема в сфере существующих онтологий заключается в том, что во многом представленные экземпляры являются не онтологиями, а таксономиями.

Многим покажется, что создание юридической онтологии превратится в длительную и трудоёмкую задачу, выполнить которую вручную не представляется возможным. Сторонники такого подхода предлагают воспользоваться альтернативным методом – автогенерацией онтологий.

Автогенерация онтологий представляет собой процесс автоматического формирования онтологии предметной области на основе обучающей выборки документов. Этот процесс состоит из нескольких этапов, на каждом из которых происходит извлечение из текста фактов или их пост-обработка для формирования какой-то части онтологии, будь то термины или объекты, концепты или же отношения между ними.

Однако данный метод для юридической онтологии не применим. Основными его недостатками являются крайне скудный набор извлекаемых связей (is-a, subclass of), которых для полноценного описания юридической картины мира в формате структурированной базы знаний явно недостаточно.

Другой проблемой является отсутствие достаточной обучающей выборки документов. В результате мы получаем разрозненные и бессистемные классы, объединённые связями одного порядка, что на практике не позволит решить поставленные задачи.

В качестве примера можно привести фрагмент из разработанной нами онтологии:

- данный фрагмент описывает процедуры подачи и принятия искового заявления к производству, описанные в статьях 125-129 Арбитражного процессуального кодекса РФ, всего ~4 страницы текста;
- функциональное предназначение – подготовка ответов на 2 вопроса: «Соответствует ли поданное исковое заявление установленным требованиям?» и «К какой категории спора относится данное исковое заявление?»;
- суммарно этот фрагмент онтологии содержит ~ 645 классов (без учёта связей), меньшее количество классов не позволяет отвечать на поставленные вопросы.

При создании онтологий для целей Legal AI мы исходим из необходимости включения такого количества классов, которого будет достаточно для отражения системы юридических знаний в полной мере.

При таком подходе свойство «reusable» применительно к онтологии не имеет значения, поскольку она направлена на решение иных задач. Необходимое количество классов зависит от глубины детализации вопросов, ответы на которые могут быть получены с помощью онтологии, а также от уровня представления знаний.

Другой пример, иллюстрирующий необходимость наличия большого количества связей и классов, – due diligence.



Due diligence – процедура составления объективного представления об объекте и субъектах правоотношений, включающая в себя оценку правовых и финансовых рисков, независимую оценку объекта (вещи, имущественные права и др.), всестороннее исследование деятельности компании, комплексную проверку её финансового состояния и положения на рынке.

Такая услуга оказывается опытными юристами-консультантами перед покупкой бизнеса, осуществлением сделок слияния и поглощения (M&A), подписанием сложных контрактов, соглашений о сотрудничестве и включает в себя полномасштабное и всестороннее исследование компании.

Обращаясь к рассмотренной ранее модели представления знаний в форме пирамиды, необходимое количество классов онтологии может быть выражено следующим образом:

- для ответа на вопрос: «Кто является директором компании?» – требуется в среднем до 30 классов (уровень data);
- для ответа на вопрос: «Какие полномочия есть у директора согласно уставу?» – требуется в среднем до 100-150 классов (уровень information);

➤ для ответа на вопрос: «Имеются ли у директора полномочия на подписание конкретной сделки?» — требуются в среднем сотни – тысячи классов (уровень knowledge);

➤ для проведения полномасштабного due diligence требуются десятки тысяч классов (уровень wisdom).

Такое количество необходимых классов для проведения due diligence обусловлено сложностью процедуры. Например, для классической услуги по проведению due diligence в отношении компании перед её приобретением юристу необходимо:

➤ определить схему владения и систему управления компанией;

➤ проанализировать хозяйственную деятельность организации и оценить её эффективность с точки зрения прибыли и наличия или отсутствия признаков банкротства;

➤ проанализировать структуру активов компании и права на них (недвижимое имущество, движимое имущество, объекты интеллектуальной собственности, иные имущественные права);

➤ оценить соответствие деятельности компании трудовому, антимонопольному, налоговому, административному законодательству;

➤ проанализировать судебные разбирательства, в которых компания принимала участие или участвует в настоящий момент и др.

Это лишь часть направлений, которые необходимо проанализировать юристу при выявлении рисков и подготовке заключения.

В зависимости от сложности и глубины практических вопросов, ответы на которые должны быть получены с помощью онтологии, зависит необходимая глубина ее детализации. Чем выше сложность вопросов, тем больше требуется классов для ответа на них.

При этом необходимо учитывать, что рост количества связей между классами экспоненциален росту количества классов онтологии.

Для сравнения в настоящий момент онтология FIBO, охватывающая только область финансовых взаимоотношений, включает в себя 3099 классов.

#### **4. Заключительные положения**

Изложенные в настоящей статье рассуждения и выводы относительно возможности практической реализации и методологии разработки решений Legal AI

были выработаны нами в течение нескольких лет исследований и практической деятельности.

Не остаётся сомнений в том, что развитие искусственного интеллекта в настоящий момент является одним из приоритетных направлений деятельности мировых технологических лидеров в том числе на государственном уровне.

6 июля 2020 года Президент РФ утвердил Поручение Правительству РФ с учётом ранее данных поручений принять исчерпывающие меры по утверждению отдельного федерального проекта «Искусственный интеллект», обеспечив необходимое финансирование, в том числе из предусмотренных на реализацию национальной программы «Цифровая экономика Российской Федерации» средств федерального бюджета.

Данный факт свидетельствует о выделении технологий искусственного интеллекта в самостоятельную ветвь национального проекта.

Мы, являясь компанией, которая занимается развитием технологий искусственного интеллекта в области юриспруденции, считаем, что такое решение является стратегически правильным.

В долгосрочной перспективе оно будет способствовать комплексному развитию всего направления в том числе при помощи государственной поддержки различных проектов, а не только отдельных субтехнологий.

Поскольку изучению вопросов развития Legal AI мы уделяем большое количество времени, у нас сформировано представление о том, какие меры будут способствовать наиболее эффективному и быстрому развитию рынка технологий искусственного интеллекта.

Важно отметить, что в общественном сознании устойчиво мнение о том, что отечественные достижения в области цифровых технологий и, в частности, в сфере искусственного интеллекта, отстают от результатов зарубежных стран (США, Китая и др.).

Однако это не так, и существующий разрыв может быть сокращён в краткосрочной перспективе.

На российском рынке сложилась уникальная ситуация, при которой созданы все необходимые условия для появления в самом ближайшем будущем прорывных отечественных решений.

Во-первых, в России существует прочная научная и практическая основа в области искусственного интеллекта. При этом многие материалы и публикации по теме находятся в открытом доступе, а в профессиональном сообществе активно осуществляется обмен опытом, достижениями и концептами. Это позволяет наглядно видеть существующие тренды (в том числе зарубежные) в области

искусственного интеллекта, понимать вектор развития и оценивать практические достижения в данной сфере.

Во-вторых, в последние годы активно развивается рынок труда и появляются специалисты в области искусственного интеллекта, машинного обучения и нейронных сетей, что позволяет создавать сильные и профессиональные команды разработчиков.

Кроме того, благодаря процессам глобализации существует возможность привлекать в команды зарубежных специалистов, обладающих богатым практическим опытом и необходимыми теоретическими знаниями.

В-третьих, весомым преимуществом для отечественного рынка программных решений на основе технологий искусственного интеллекта является уникальность русского языка. Это не позволяет зарубежным IT-компаниям, являющимся лидерами в области цифровых технологий, создавать решения для отечественного рынка и конкурировать с российскими разработчиками, уступающим им по масштабам и финансовым возможностям.

В совокупности данные факторы создают все условия для самостоятельного развития технологий искусственного интеллекта в России силами отечественных разработчиков, которые при правильном подходе имеют все шансы на успех.



## Статья 3

### Искусственный интеллект в области юриспруденции.

#### Введение

Мы продолжаем цикл статей на тему юридического искусственного интеллекта, аспектов его разработки и перспектив практического применения на отечественном рынке.

В предыдущих публикациях мы неоднократно говорили, что, по нашему мнению, разработка Legal AI может быть обеспечена с помощью создания и применения нового семантического блока, включающего в себя:

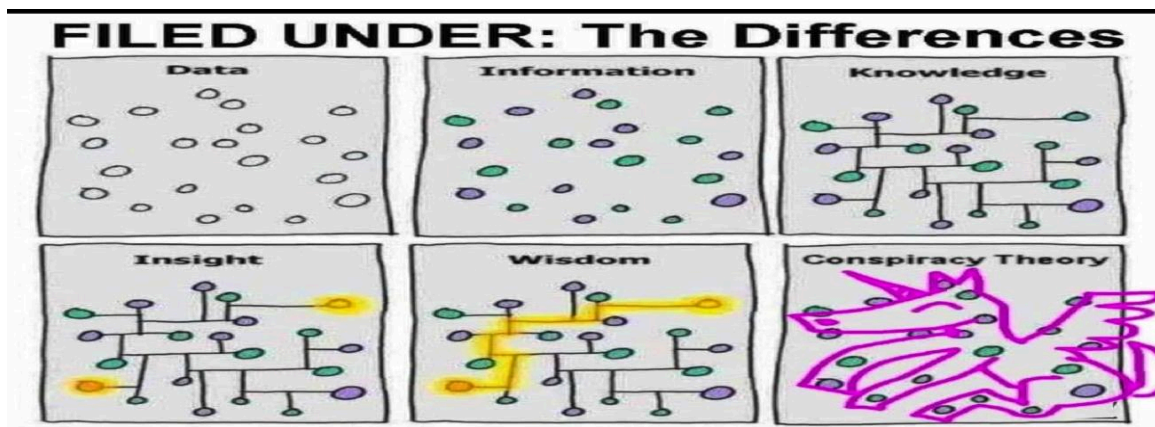
- инструменты лингвистического анализа текстов на естественном языке;
- структурированную модель юридических знаний (графы знаний и онтологии);
- предобученные нейронные сети.

В первой статье мы детально исследовали существующие инструменты процессинга русскоязычного текста.

Во второй статье мы рассмотрели подходы к созданию продуктов на основе искусственного интеллекта, а также вопросы взаимодействия специалистов в области IT и юриспруденции.

В настоящей статье мы предлагаем погрузиться в тему онтологий и ответить на следующие вопросы:

1. Какова роль онтологий в процессе создания искусственного интеллекта?
2. Почему существующие онтологии в области права неприменимы для Legal AI, несмотря на многолетние попытки зарубежных специалистов структурировать юридические знания?
3. Какими свойствами должны обладать онтологии для Legal AI, чтобы решать практические задачи?



## 1. Логическая интерпретация данных

### 1.1. Уровни описания знаний

Применение комплекса инструментов процессинга текста на естественном языке, как мы отмечали ранее, позволяет решить первую задачу на пути к Legal AI, а именно – добиться глубокого уровня детализации смыслового содержания юридических документов при помощи группы методов лингвистического анализа (морфология, синтаксис, семантика), машинного обучения и др. Представим, что «завтра» данная задача для русского языка решена на 99,99%.

Следующий важный шаг – это интерпретация результатов с точки зрения предметной области (reasoning). Именно на данном этапе происходит соотношение полученных данных с юридической картиной мира и определение правового содержания того или иного документа.

Например, традиционно в деловой практике при составлении любых гражданско-правовых договоров одним из первых разделов является «Предмет договора», который закрепляет сущность отношений между сторонами. Иными словами, в предмете договора содержатся ответы на вопросы: «По поводу чего заключён договор, и каким образом стороны воздействуют на определённый объект?». Так, в договоре купли-продажи предмет обычно сформулирован следующим образом: «В соответствии с условиями настоящего договора продавец обязуется передать товар в собственность покупателя, а покупатель обязуется принять товар и оплатить его стоимость».

На основе формулировки предмета договора юристы чаще всего определяют его правовую природу, что позволяет правильно квалифицировать отношения между сторонами и применить нужные нормы права, соответствующие характеру правоотношений.

Безусловно, это простой и наиболее распространённый пример.

Для того, чтобы по такой формулировке идентифицировать договор купли-продажи, экспертных юридических знаний, как правило, не требуется.

Но в реальной юридической практике мы чаще сталкиваемся со сложными экономическими взаимоотношениями сторон, которые выходят далеко за рамки простых договоров (купли-продажи) и требуют глубокой работы по толкованию содержания контракта, чтобы дать ему правовую оценку.

Однако от сложности конкретной ситуации не зависит общий методологический подход. Человек-юрист читает текст договора или иного документа, идентифицирует в его содержании ключевые юридические факты и смыслы, структурирует их в соответствии с известными ему юридическими концептами и, наконец, применяет системные знания и опыт для формирования логических суждений и выводов.

При этом очень часто данные процессы происходят в сознании юриста неразрывно и параллельно, поскольку любой человек воспринимает и обрабатывает входящую информацию под призмой собственных профессиональных знаний.

Когда мы делегируем программным инструментам любую работу, связанную с логической обработкой данных, необходимо, в первую очередь, сформировать единую картину мира, которая станет общей базой знаний, доступной для восприятия человеком и программными инструментами.

В случае с программными решениями для юристов, целью которых является логическая обработка данных, требуется концептуальное описание юридической картины мира, закономерностей и взаимосвязей правовой системы, глубина которых определяется спектром решаемых задач.

Чем сложнее задачи, тем более глубокая и масштабная оцифровка юридических знаний необходима для достижения нужного результата.

Говоря о формализации знаний в контексте data science, нельзя не упомянуть описанную Judea Pearl концепцию из 3 уровней: Association, Causation и Counterfactuals (*его книга на эту тему настоятельно рекомендуется к прочтению*).

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y x)$	Seeing	What is? How would seeing $X$ change my belief in $Y$ ?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y do(x), z)$	Doing	What if? What if I do $X$ ?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x x', y')$	Imagining, Retrospection	Why? Was it $X$ that caused $Y$ ? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Уровень Association является базовым и наиболее распространённым методом в машинном обучении. Он предполагает генерацию суждений на основе статистики, полученной по итогам обработки большого объёма исходных данных. Особенность данного уровня заключается в том, что получаемый конечный результат с высокой долей вероятности будет верным, однако логическое объяснение такого результата будет отсутствовать.

Следующий уровень – Causation. На данном уровне присутствует поверхностная причинно-следственная связь, которая позволяет приходить к определённому результату на основе ряда отличительных признаков.

Например, из десятков фотографий животных можно идентифицировать изображение, на котором изображена кошка, с помощью обнаружения уникальных анатомических признаков, присущих именно кошке. Уровень Causation существенно превосходит предыдущий уровень по возможностям прогнозирования и анализа, однако выявляемая причинно-следственная связь является прямой.

Для примера: на уровне Causation мы можем утверждать, что рак лёгких способен привести к летальному исходу. Но на уровне Causation мы не сможем с уверенностью сказать, что летальный исход явился следствием именно этой вредной привычки, а не какого-либо иного фактора.

И, наконец, третий уровень – Counterfactuals. Данный уровень включает в себя возможности глубоких причинно-следственных рассуждений и способен давать ответы на вопросы, требующие ретроспективного анализа.

Возвращаясь, к примеру с курением, с помощью данного уровня можно не только утверждать, что курение может привести к летальному исходу, но и формировать рекомендации по снижению рисков заболеваемости раком лёгких, игнорировать незначительные факторы и др.

Как мы уже отмечали ранее, наиболее распространённым уровнем знаний в машинном обучении является Association.

По своей сути все методы достижения результата сводятся к поиску идеального математического уравнения / функции, позволяющей наиболее точно ответить на необходимый вопрос.

Однако, с точки зрения юриста-профессионала, такая «подгонка под результат» вряд ли когда-нибудь станет общепринятым инструментом.

В юриспруденции идеальным уровнем, на котором должно оперировать ПО, является Counterfactuals, который предполагает использование полноценной и детализированной картины мира, позволяющей находить ответы на самые разнообразные вопросы и прогнозировать результат на основе произвольных вводных данных. Именно на данном уровне мы можем говорить о возможности формировать суждения и умозаключения в какой-либо предметной области, соответствующие функционалу цифрового юриста.

## 1.2. Построение суждений и умозаключений

Построение суждений и умозаключений в отношении входных данных на основе юридической логики является ключевой задачей Legal AI.

В процессе разработки и создания инструментов, обеспечивающих решение данной задачи, как мы отмечали в предыдущих публикациях, необходимо руководствоваться «first principles» и подходить к её реализации с самых фундаментальных начал.

Одной из основных возможностей и назначений онтологий как семантических моделей представления знаний предметной области является возможность генерации и вывода новых знаний на основе взаимосвязей между имеющимися концептами (reasoning). Результаты данного процесса включают в себя два элемента: суждение и умозаключение. Их отличия берут свое начало в философии и логике, которые изучают процессы мышления и познания. Так, суждением является мысль, в которой посредством связи понятий утверждается (отрицается) что-либо о чем-либо, а умозаключение – более высокий уровень логического опосредования, чем суждение, в ходе которого из сопоставления ряда суждений выводится новое суждение.

Возвращаясь к рассматриваемой проблематике, можно привести классический пример суждения: «S есть P». Генерация подобных суждений может быть обеспечена при помощи набора лингвистических инструментов с применением семантики и графа знаний.

В результате мы можем говорить о том, что задача процессинга текста сводится к трансформации предложений на естественном языке в набор подобного рода суждений.

Практический пример реализации данного инструментария может быть проиллюстрирован с помощью анализа вводной части («шапки») искового заявления, в которой традиционно указываются следующие сведения:

- наименование и адрес суда, в который адресован данный иск;
- наименование, идентификаторы (ОГРН/ОГРИП, ИНН, КПП и др.) истца и его адрес;
- наименование, идентификаторы (ОГРН/ОГРИП, ИНН, КПП и др.) ответчика и его адрес и др.

Например, в части сведений об истце/ответчике шапка искового заявления обычно содержит в себе его наименование и адрес в формате: «Истец: ООО «Ромашка», г. Москва, ул. Тверская, д. 1, офис 1».

Из данного фрагмента текста может быть выделен набор суждений, которые изначально не имеют прямой связи между собой.

Такие связи подразумеваются, и в результате их восстановления мы можем с помощью инструментов лингвистики и NER получить следующие суждения:

- «ООО «Ромашка»» есть «Организация» типа «Общество с ограниченной ответственностью»;
- «г. Москва, ул. Тверская, д. 1, офис 1» есть «Адрес».

При этом подразумевается наличие ещё одной связи, следующей из данного фрагмента: «ООО «Ромашка» имеет «Адрес» равный «г. Москва, ул. Тверская, д. 1, офис 1». Данная связь является примером умозаключения, которые мы получаем путём сопоставления двух суждений. Это и есть результат процесса reasoning'a.

Для того, чтобы генерировать умозаключения, необходимо иметь набор формализованных знаний, на основе которых становится возможным сопоставлять суждения и выводить новые знания о конкретных субъектах и объектах отношений.

В нашем примере таким знанием является правило о том, что «Организация» типа «Общество с ограниченной ответственностью» имеет «Адрес».

Очевидный вопрос в данной ситуации: «Каким образом формировать данные формализованные знания?».

Здесь нам на помощь приходит дескрипционная логика и объектно-ориентированное программирование.

### 1.3. Дескрипционная логика и объектно-ориентированное программирование

Дескрипционная логика (описательная логика, логика концептов) – язык представления знаний, позволяющий описывать понятия предметной области в недвусмысленном, формализованном виде, организованный по типу языков математической логики. Дескрипционные логики сочетают, с одной стороны, богатые выразительные возможности, а с другой – хорошие вычислительные свойства, такие как разрешимость и относительно невысокая вычислительная сложность основных логических проблем, что делает возможным их применение на практике, обеспечивая компромисс между выразительностью и разрешимостью.

Современное название семейство дескрипционных логик получило в 1980-е годы. В то время они изучались как расширения теорий фреймовых структур и семантических сетей механизмами формальной логики.

В 2000-е годы дескрипционные логики получили применение в рамках концепции семантической паутины, где их предлагалось использовать при построении онтологий. Язык OWL разрабатывается как язык, на котором можно формулировать и публиковать так называемые сетевые онтологии – формально записанные утверждения о понятиях и объектах некоторой предметной области.

Одно из требований к таким онтологиям заключается в том, чтобы содержащиеся в них знания были доступны для машинной обработки, в частности, для автоматизированного логического вывода новых знаний из уже имеющихся. Для этого требуется, чтобы язык, на котором формулируются онтологии, имел точную семантику, а соответствующие логические проблемы были разрешимы (и имели практически допустимую вычислительную сложность).

Кроме того, желательно, чтобы такой язык имел довольно большую выразительную силу, пригодную для формулировки на нём практически значимых фактов.

Дескрипционные логики обладают такими свойствами, и по этой причине они были выбраны в качестве логической основы для языка OWL. OWL базируется на языке разметки XML, поэтому можно сказать, что OWL является результатом преобразования некоторых дескрипционных логик с использованием синтаксиса XML.

Таким образом, мы подошли к исследованию онтологий как формы структурирования информации. Подобные модели представления знаний нашли своё практическое применение в широких областях. Самое очевидное и первостепенное применение – использование онтологий в качестве источника данных для компьютерных приложений (для информационного поиска, анализа текстов, извлечения знаний и др.), поскольку онтологии позволяют более

эффективно обрабатывать сложную и разнообразную информацию. Этот способ представления знаний позволяет распознавать те семантические отличия, которые являются само собой разумеющимися для людей, но не известны программным инструментам.

На современном этапе своего развития онтологии и графы знаний становятся инструментами выражения знаний и опыта предметной области, в том числе для решений на базе технологий искусственного интеллекта в самых различных сферах, в том числе и в области юриспруденции.

Говоря о необходимости создания юридических онтологий и графов, мы не ставим перед собой задачу по переводу в машиночитаемый формат норм права из текстов законов, иных нормативно-правовых актов, а также судебных решений. Подобные идеи возникали в научном сообществе, однако, по нашему мнению, они не имеют практического смысла.

Юриспруденция является крайне сложной и фрагментарной областью знаний несмотря на то, что имеется общепринятая в РФ система права, которая позволяет выделить определённые отрасли, подотрасли, институты и др.

На практике реальные кейсы очень часто лежат на пересечении многих отраслей права, имеющих в том числе противоположное регулирование.

В теории права даже существуют такие понятия, как пробелы правового регулирования (наличие отношений, не имеющих какого-либо правового регулирования), коллизии права (случай, при котором одна ситуация подвергнута противоположному регулированию), аналогия права и др.

Кроме того, нормативно-правовые акты задают общий вектор правового регулирования, часто определяя лишь основные начала и его принципы, оставляя диспозитивным субъектам отношений право самостоятельно согласовать для себя правила поведения в установленных пределах.

Например, в отношении отдельных видов договоров ГК РФ закрепляет определённые существенные условия, без согласования которых договор считается незаключённым. Как правило, это предмет договора, срок и/или цена (точный набор существенных условий зависит от конкретного вида обязательства). Они выступают минимально необходимым числом условий для заключения договора, но в деловой практике редки случаи, когда договор ограничивается только существенными условиями, упомянутыми в законе.

Очень часто стороны самостоятельно формулируют подходящие для них условия сотрудничества, которые в законе не упомянуты.

Кроме того, в гражданском праве существуют такие категории, как непоименованные договоры (соглашения, не предусмотренные ГК РФ в качестве

самостоятельного вида договора), регулирование которых осуществляется по аналогии закона/права.

Помимо этого, тексты законов определяют только правило поведения. Возвращаясь, к примеру про существенные условия договора, важно отметить, что ГК РФ требует наличия данных условий в договоре, но не определяет порядок их согласования, рекомендуемые формулировки того или иного условия и др. Все это свидетельствует о том, что тексты нормативно-правовых актов и судебных решений охватывают, по нашей оценке, не более 20-30 % реальной юридической практики. Многие вопросы правоприменения лежат за их рамками и формируются в формате обычаев делового оборота, поэтому трансформация законов и судебных решений в машиночитаемый формат не позволит в полном объеме воспроизвести юридическую картину мира.

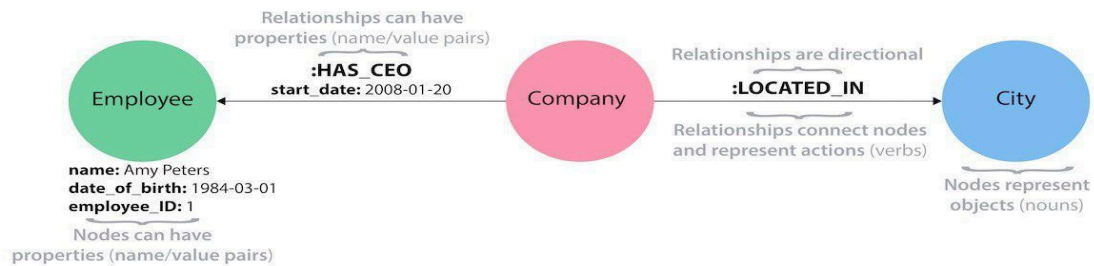
Отдельная проблема данного подхода – неоднозначность и сложность юридического языка. В предыдущей публикации мы рассказывали об эксперименте 1986 года по разбору нормативного акта (The British Nationality Act) с применением логики языка Prolog.

Суть эксперимента заключалась в попытке разложения текста данного закона на структурные единицы (смысловые триплеты), что и было осуществлено. Авторы эксперимента тогда отмечали, что большой проблемой для самих правоприменителей (юристов, адвокатов, судей и др.) является неоднозначность и двусмысленность законов, которые создают дополнительные проблемы при попытке их толкования в процессе применения. Сложность профессиональной лексики, проблемы в юридической технике авторов законов и многое другое не всегда позволяют однозначно истолковать тот или иной нормативно-правовой акт, приводя к разному пониманию одной и той же нормы.

Данная проблема актуальна для всех юрисдикций, в том числе и для России. Достаточно вспомнить сложившуюся практику применения новых норм. После принятия и вступления в силу новых положений законодательства часто требуется значительное количество времени для выработки практики их применения, в том числе нередки случаи, когда за принятым законом следуют разъяснения ВС РФ в виде Постановлений Пленума, которые объясняют, как правильно интерпретировать те или иные положения. Все это говорит о том, что оцифровка текстов нормативно-правовых актов и судебных решений не позволит в полной мере воссоздать юридическую картину миру и выработать базу правовых знаний.

Отдельно остановимся на ограничениях и трудностях структурирования юридических знаний в формате онтологий.

Язык OWL обладает рядом ограничений, которые порождают сложности при попытках отражения отдельных видов отношений, существующих в объективной реальности.



К таким отношениям могут быть отнесены почти все связи, которые на естественном языке выражаются с помощью причастий и деепричастий, а также отношения, которые длились в течение определённого периода времени. Например, утверждение: «Генеральный директор, назначенный решением собрания акционеров, осуществлял полномочия в период с 01.09.2015 г. по 01.09.2019 г.» на практике крайне трудно смоделировать в OWL, поскольку связи (object properties) не предполагают наличие собственных свойств.

Существующие в OWL подходы к решению данной проблемы создают приводят к практически неконтролируемому росту классов, в результате чего:

- онтологии становится крайне сложно интерпретировать и поддерживать;
- операции с классами и инстансами (instances) становятся сложновычислимыми.

Данная проблема характерна в принципе для всего объектно-ориентированного программирования, однако в последнее время появляются перспективные проекты и решения в этой области.

Например, сообществом W3C в 2017 г. принят стандарт Shapes Constraint Language (SHACL), который позиционируется в качестве альтернативы RDFS/OWL и обладает более широкими возможностями по описанию свойств и характеристик объектов.

Кроме того, потенциальным решением является переход к нереляционным базам данных (графовым системам), которые обладают гораздо большей выразительностью при описании классов и связей.

## 2. Графы знаний и онтологии

### 2.1. Обзор существующих юридических онтологий

В рамках предыдущей публикации мы исследовали отдельные вопросы, связанные с онтологиями. Тогда мы писали, что онтология является унифицированной и структурированной базой знаний предметной области, представляющей собой объективное семантическое отражение картины мира в виде набора связанных между собой терминов, иерархически записанных в формате классов, подклассов и связей (Relationships) между ними, что позволяет соотносить имеющиеся данные между собой с точки зрения экспертной логики.

Также мы отмечали, что среди способов систематизации знаний онтологии предшествовала таксономия (любая структура знаний в виде иерархически соотносящихся сущностей с связями одного типа – «subclass of»), а органическим развитием онтологических систем представления знаний является граф. Граф структурно включает в себя графовые хранилища семантических метаданных и онтологий, которые в данном случае выступают в роли полуструктурированной модели предметной области, являясь ядром графа знаний.

В результате такого способа обеспечивается возможность решения интеллектуальных задач с помощью постоянной циркуляции данных за счет применения методов машинного обучения.

Вопросы о необходимости структурирования и систематизации знаний в различных предметных областях впервые начали возникать еще в XX веке.

Активное применение онтологий в качестве модели такой систематизации началось за рубежом в 1990-х годах, и на тех этапах был создан ряд обширных онтологий, включающих несколько тысяч терминов (OMEGA, SUMO, DOLCE и другие).

В начале 2000-х годов были выработаны единые стандарты построения онтологий. Так, в 2004 г. объединенным консорциумом W3C была опубликована первая рекомендация и спецификации языка OWL 1.0, который получил своё развитие до версий 1.1 и 2.0. Следует отметить, что до принятия спецификаций языка OWL их развитие происходило с применением других языков представления онтологий (KIF, Ontolingua, XOL и др.). В России единый подход не был выработан, поэтому развитие онтологических систем знаний происходит менее активно (как и их практическое применение). Первые онтологии создавались в основном в рамках научных исследований. Позднее онтологии стали получать своё распространение в

самых различных прикладных областях (медицина, биология, информатика, финансы, бизнес-сфера и др.).

Не стала исключением и юриспруденция. Первые функциональные онтологии в области права появлялись еще в 1990-х годах (Functional Ontology of Law by Valente et al. (1994), Frame-based ontology by van Kralingen (1997) и др.).

В настоящее время наиболее распространенными в мире являются следующие юридические онтологии: LKIF, UFO, FIBO, FBO и Legal Rule ML. Мы решили детально исследовать и сравнить данные онтологии между собой (подробнее в Разделах 2.1.1-2.1.6).

Прежде чем перейти к анализу конкретных онтологий, необходимо определить критерии сравнения и оценки их практической применимости.

На основе длительных теоретических исследований и практической деятельности по моделированию онтологий и графов для оцифровки отдельных правовых знаний мы выработали собственное видение идеальной модели представления юридической картины мира и требований к ним.

В первую очередь, необходимо понимать, что создать универсальную и общеприменимую онтологию в области права невозможно в силу различия государственно-правовых систем и применяемого языка.

Мы детально рассматривали данные причины в предыдущей публикации, в которой приводили пример того, что даже внутри одной правовой семьи существуют значительные и фундаментальные отличия в понимании одних и тех же правовых институтов и явлений.

Кроме того, говоря об уровне детализации онтологий, мы отмечали, что существующие модели знаний не отличаются глубоким уровнем проработки, поскольку они создавались преимущественно для целей обмена знаниями и информацией, что требует использования высокой степени абстракции, а потому они не применимы для целей практической юриспруденции, поскольку используемых концептов будет недостаточно для отражения правоотношений.

С учётом данных обстоятельств в качестве критериев сравнения рассматриваемых далее 5 (пяти) онтологий мы предлагаем использовать следующие:

1. уровень детализации (количество понятий и концептов);
2. возможность структурирования концептов из правовой системы РФ;
3. возможность и необходимость доработки онтологии для обеспечения совместимости с правовой системой РФ;
4. логика и удобство использования;

5. ширина охвата правоотношений;
6. необходимость модификации при изменении законодательства.

### 2.1.1. FIBO

Financial Industry Business Ontology (FIBO) является одной из наиболее распространенных онтологий в финансовой сфере. Она разрабатывается международным сообществом под управлением консорциума OMG, который занимается объектно-ориентированными технологиями и стандартами с 1989 года.

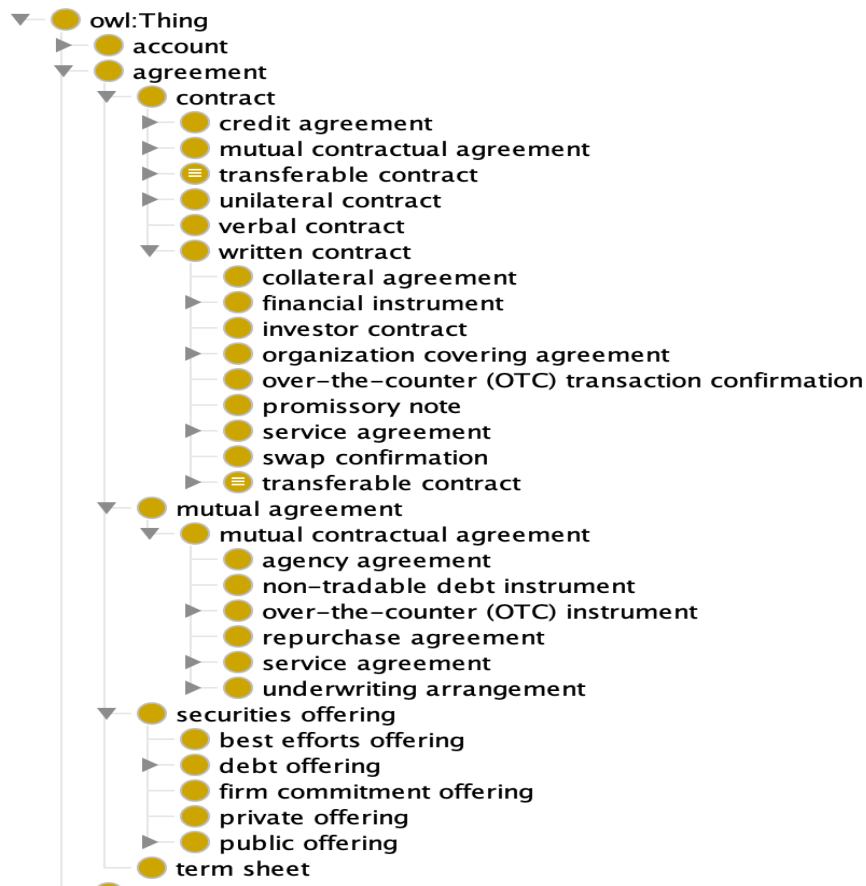
Стандарты FIBO применяются многими налоговыми органами различных государств в том числе в рамках автоматического обмена информацией (сведения о бенефициарах и др.).

Данная онтология в контексте анализа моделей представления юридических знаний имеет неоднозначное значение.

С одной стороны, FIBO описывает финансовый рынок и отношения между его участниками (финансовые сделки/операции, финансовые активы и др.), не имея в качестве своей основной цели структурировать правовые категории и стать стандартом юридической онтологии.

С другой стороны, в силу наличия значительных пересечений финансовой сферы и права описывать финансовые отношения без использования базовых правовых категорий невозможно, поэтому FIBO включает в себя правовой блок терминов, минимально необходимых для структурирования финансовой области знаний.

Так, например, для построения финансовой онтологии невозможно не использовать такие классические правовые понятия, как юридические лица, ценные бумаги, сделки и многое другое.



Однако мы отмечали в предыдущей статье, что рассматривать FIBO в качестве юридической онтологии неправильно.

Во-первых, вследствие различия глубины проработки вопросов финансов и права онтология FIBO малоприменима для решения задач в сфере Legal AI, а, во-вторых, даже имеющиеся в ней юридические концепты построены на зарубежных правовых системах, которые не адаптированы для российского права.

Но несмотря на данные факты, онтология FIBO обрела крайне широкое распространение и практическое применение в бизнес-сфере, а также считается одним из лучших примеров онтологий в целом, поэтому мы решили детально проанализировать ее состав и особенности.

FIBO совокупно состоит из 3099 классов. Понятия онтологии FIBO представлены двумя способами: формальным описанием понятий и их взаимосвязей на языке OWL, а также их описанием на естественном языке с использованием толковых словарей финансовой отрасли.

Формальное описание понятий на языке OWL выполнено с применением широко используемого в семантическом моделировании редактора онтологий Protege. Предполагается, что онтология FIBO должна стать общим языком для финансовой индустрии, поддерживающим автоматизацию бизнес-процессов. Она предназначена для использования разработчиками программного обеспечения,

бизнес-аналитиками и другими участниками сферы финансов. Бизнес-термины и определения, описанные в FIBO, могут быть использованы в качестве эталонной модели, с которой финансовые организации могут связывать свои собственные (локальные) модели. Появляется возможность создавать логические модели данных, которые получают из FIBO свою формальную семантику.

В основе FIBO лежит онтология верхнего уровня, которая включает категории, называемые «разделами».

Разделы включают описание различных типов базовых сущностей (например, таких, как независимые/связанные, реальные/абстрактные и т.п.).

Онтологию верхнего уровня использует базовый модуль FIBO Foundations, который содержит большой набор подмодулей.

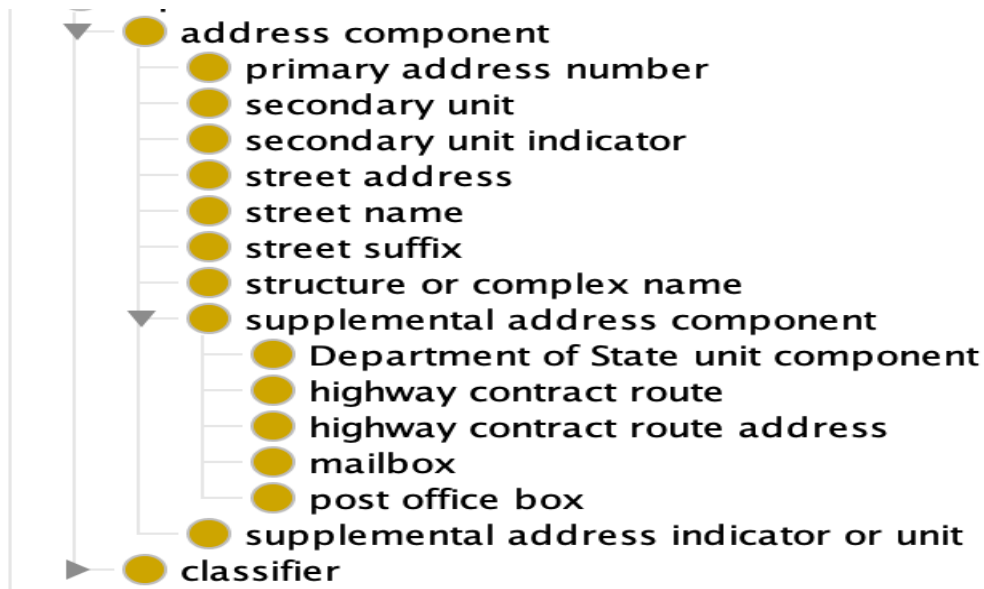
А на основе FIBO Foundations уже разрабатываются такие модули, как FIBO Business entities (бизнес-сущности) и FIBO Indices and Indicators (индексы и показатели).

Такая модульная формальная модель позволяет отдельно описать, использовать и при необходимости дополнять группы понятий, содержащихся в отдельных модулях.

FIBO Foundations включает в себя набор базовых понятий, которые в дальнейшем используются для описания соответствующих им терминов финансовой отрасли. В этом модуле описываются общие понятия, которые не являются уникальными для финансовой индустрии, но требуются для описания понятий, связанных с финансами, в том числе набор базовых юридических категорий (договоры и сделки, организации и др.).

В некоторых случаях в FIBO Foundations используются понятия-заместители («Proxu» concept) для выполнения ссылок на понятия, описанные в других разделах. Например, подобные заместители используются для таких понятий, как адрес (address) или страна (country).

Это связано с двумя причинами: понятия, используемые в финансовой отрасли, обычно имеют более специфичный смысл, чем соответствующие им более общие, нефинансовые понятия.



Примерами таких понятий являются контракты или транзакции, которые включены в FIBO Foundations для того, чтобы их можно было уточнить в других спецификациях онтологий.

Вторая причина использования понятий-заместителей заключается в том, что свойства некоторых понятий финансовой отрасли часто должны использоваться во взаимосвязи с другими, не финансовыми понятиями (например, страны, юрисдикции, адреса и т.п.).

Такие понятия включены в FIBO Foundations, чтобы на них можно было сослаться в других спецификациях FIBO.

Раздел	Содержание
Utilities (утилиты)	Описывает часто используемые типы данных, без использования их программных обозначений (например, <i>URI</i> ) в понятном бизнес-аудитории виде
Relations (отношения)	Определяет набор отношений, которые используются для описания новых отношений между конкретными понятиями в других разделах <i>FIBO</i>
Goals and Objectives (цели и задачи)	Включает цели и задачи людей или организаций. Понятие «цели» является основным для определения деятельности организации, а понятие «задачи» необходимо для описания бизнес-планов их работы
Parties (стороны)	Описывает такие понятия, как «участник» и «агент», а также независимые роли
Agents and people (агенты и люди)	Содержит понятия, относящиеся к типам автономных единиц (сущностей), способных управлять своим поведением
Places (места)	Описывает реальные или виртуальные места и их адреса
Organizations (организации)	Определяет типы и особенности организаций: формальные, неформальные, законные или незаконные и т.д.
Agreements (соглашения)	Описывает соглашения между сторонами и формализующие их договоры (контракты). Данный раздел включает определения письменных и устных договоров и их свойства
Law (закон)	Содержит понятия, определяющие виды источники права
Ownership and Control (собственность и управление)	Включает понятия, определяющие значения собственности, активов и владельца, типов управления
Accounting (бухгалтерский учет)	Состоит из общих концепций бухгалтерского учета
Dates and Time (дата и время)	Содержит понятия, описывающие даты и времена с учетом особенностей финансовой отрасли

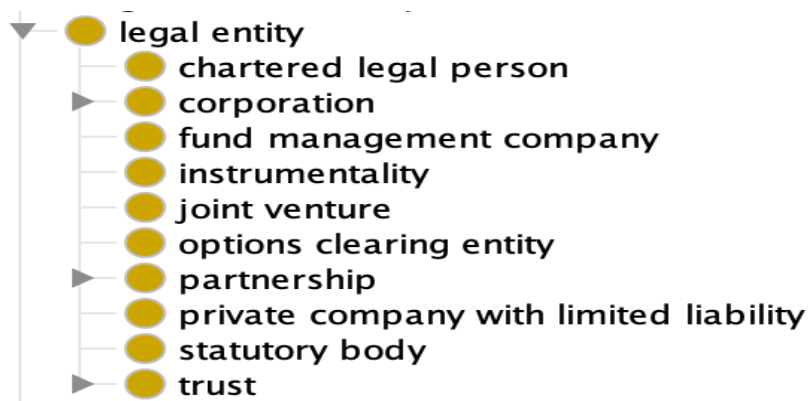
Раздел FIBO Business entities не входит в состав FIBO Foundations и создан на его основе.

Он представляет собой модель бизнес-концепций, которые представлены в терминах финансовой индустрии, используемых в нормативных и финансовых документах организаций.

В данном модуле описан набор терминов, связанных с хозяйствующими субъектами (business entities), в том числе юридическими лицами различных организационно-правовых форм, отношениями управления и собственности между ними.

Также FIBO Business entities содержит подробные требования к формированию диаграмм и отчётов, предназначенных для использования бизнес-экспертами.

FIBO Business entities включает набор подмодулей, позволяющих описывать юридические лица, корпорации, партнёрства, трасты, собственность и контроль, функционально определённые хозяйствующие субъекты и др.



Подводя итог обзору FIBO, отметим следующие выводы. FIBO включает в себя 3099 понятий, описывающих финансовую сферу и взаимоотношения её субъектов, не все из которых имеют отношение к юридической сфере.

Среди правовых категорий, используемых в данной онтологии, применяются только базовые понятия субъектов финансовых отношений (организации и физические лица) и объектов финансового рынка (акции и иные ценные бумаги).

При этом в онтологии в юридической части воспроизведена система англо-американского права, которая не подходит для применения в РФ.

Модульная структура онтологии положительно сказывается на возможностях её модификации. В результате отметим, что FIBO, хотя и является примером детализированной отраслевой онтологии, для решения практических задач в области

Legal AI в России не подходит ввиду того, что унификация знаний в области бизнеса и финансов не охватывает в необходимой мере юридические концепты.

## 2.1.2. LKIF

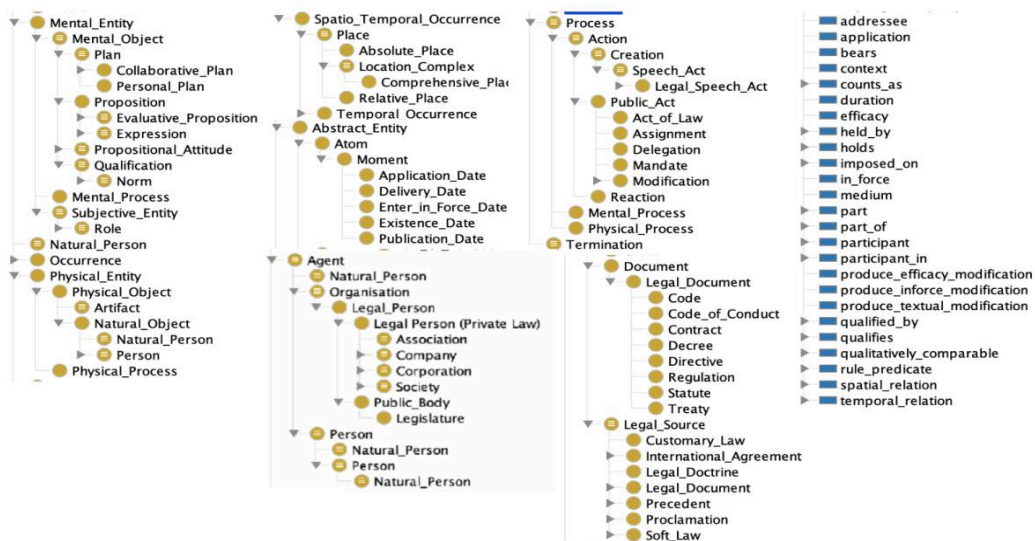
Одной из наиболее объемных юридических онтологий является LKIF (Legal Knowledge Interchange Format), созданная в рамках проекта Estrella.

В LKIF правовые концепции моделируются на основе объективной реальности и основываются на объектах и процессах реального мира.

По замыслу создателей данной онтологии на практике она должна способствовать осуществлению двух основных функций: выполнять перевод между правовыми базами знаний, написанными в разных форматах, и представлять юридические данные в формализованном формате для разработки правовых систем знаний.

В действительности LKIF является не единой онтологией, а представляет собой набор большого количества онтологий, которые разделены по модулям и подмодулям. Модули (подмодули) включают в себя наборы совместно используемых онтологий.

Так, LKIF включает в себя когда-то самостоятельную объёмную онтологию LRI-core ontology. LRI-core представляет собой базовую онтологию, содержащую основные понятия, общие для всех правовых систем, такие как норма права, мораль, контракт, роль, юридическая ответственность и др.



LKIF представляет собой библиотеку юридических онтологий, состоящую из 15 модулей, каждый из которых описывает набор тесно связанных понятий как из области права, так и общих концептов, необходимых для отражения объективной реальности. Среди данных модулей можно выделить следующие:

1. Верхний модуль – описывает такие основные понятия и концепции, как «Ментальный объект» (Mental entity), «Абстрактный объект» (Abstract entity), «Событие» (Occurrence), «Физический объект» (Physical entity);
2. Модуль связей – описывает отношения между частью и целым в различных смыслах. Включает такие отношения, как: «является компонентом» (is component), «является частью» (is part of), «является членом» (member of) и другие;
3. Модуль места – включает понятия и отношения, описывающие пространственные связи между объектами реального мира;
4. Модуль времени – описывает временные характеристики и отношения. Содержит концепции «момент» (самая малая единица времени) и «интервал» (может содержать моменты, а также другие интервалы). Основные типы отношений: «произошло до» (before), «завершилось» (finish), «между» (between) и другие;
5. Модуль процессов – содержит описания процессов изменения объектов реального мира. Выделено три типа изменений: начало, продолжение, прекращение;
6. Модуль агентов – субъекты, совершающие действия и взаимодействующие друг с другом. Типы агентов: индивиды, организации;
7. Модуль действий – описывает возможные реакции и действия агентов. Действия могут осуществляться с помощью предметов;
8. Модуль ролей – описывает социальную структуру, а также допустимое и недопустимое поведение агентов.
9. Модуль описания классов – описывает утверждения, характеристики других классов и выступает основой для описания содержимого юридических документов;
10. Модуль юридических действий – дополнительный модуль к действиям, определяющий действия исполнителей в зависимости от их определения в каком-либо нормативном документе;
11. Модуль описания юридических норм – расширяет модуль описания высказываний и определяет различные виды юридических документов: контракт (Contract), директива (Directive), кодекс (Code) и другие.

Отдельно остановимся на описании юридических концептов (Legal Concepts). Legal Concepts является основным блоком онтологии LKIF и представлен в виде трех модулей: Legal role, Legal action и Legal norm.

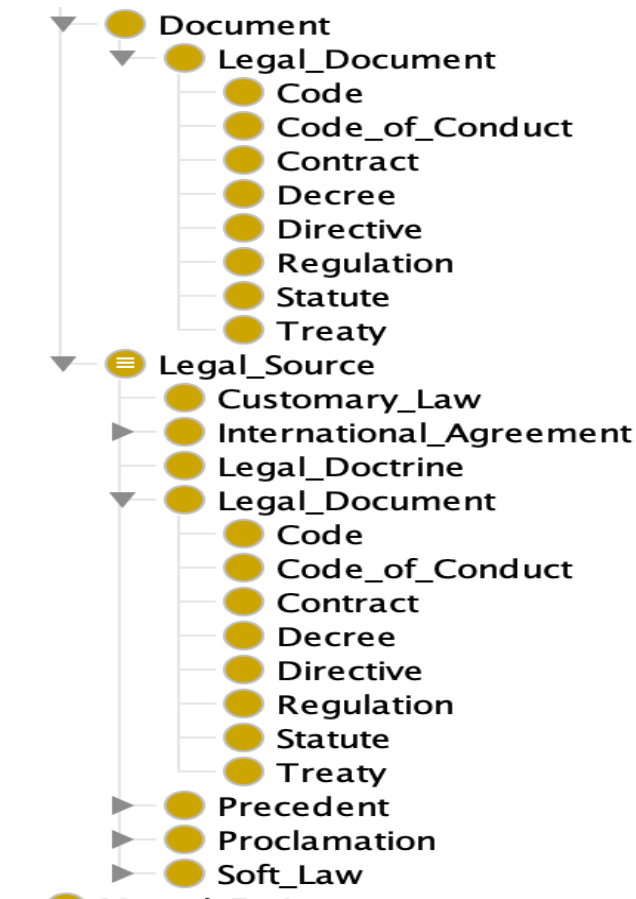


Legal role направлен на расширение общего модуля Role, описывающего возможные роли активных субъектов материального мира, путём выделения специальных правовых субъектов, имеющих особый юридический статус.

Он включает в себя такие классы, как государственные органы, юридические лица, физические лица, агенты и др.

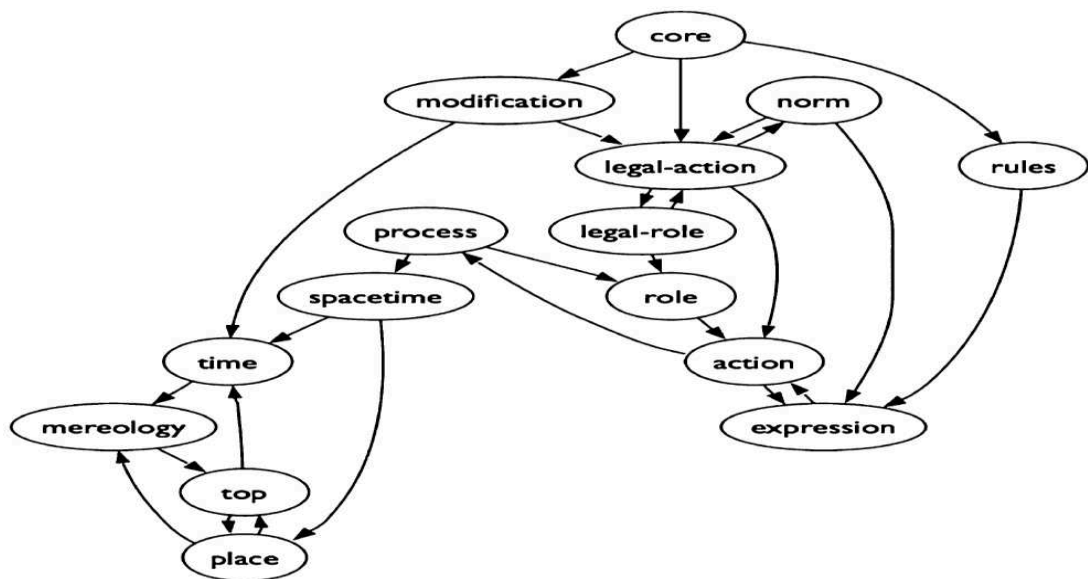


Legal action направлен на расширение модуля Action и включает в себя описание действий, осуществляемых классами, включенными в модуль Legal role.



Модуль Legal norm является расширением прежде всего модуля выражений, где нормы определяются как квалификации.

Модуль содержит в себе классы типа: законы, подзаконные акты, постановления и др., тем самым детализируя возможные источники права.



Несмотря на обширность модулей, входящих в LKIF, она является базовой и верхнеуровневой юридической онтологией, которая при этом адаптирована под систему англо-саксонской правовой семьи.

В результате такая модель представления знаний имеет некоторую универсальность и способна выполнять задачи, для решения которых она создавалась, однако применение её в качестве полноценной юридической базы знаний невозможно, поскольку она имеет высокий уровень абстракции.

Важно отметить, что LKIF исследовалась отечественными учёными (П.А. Ломов и А.Г. Олейник), которые предпринимали попытки структурирования текста отечественного нормативного акта на основе данной онтологии.

В качестве такого акта использовалась «Стратегия социально-экономического развития Мурманской области до 2020 года и на период до 2025 года».

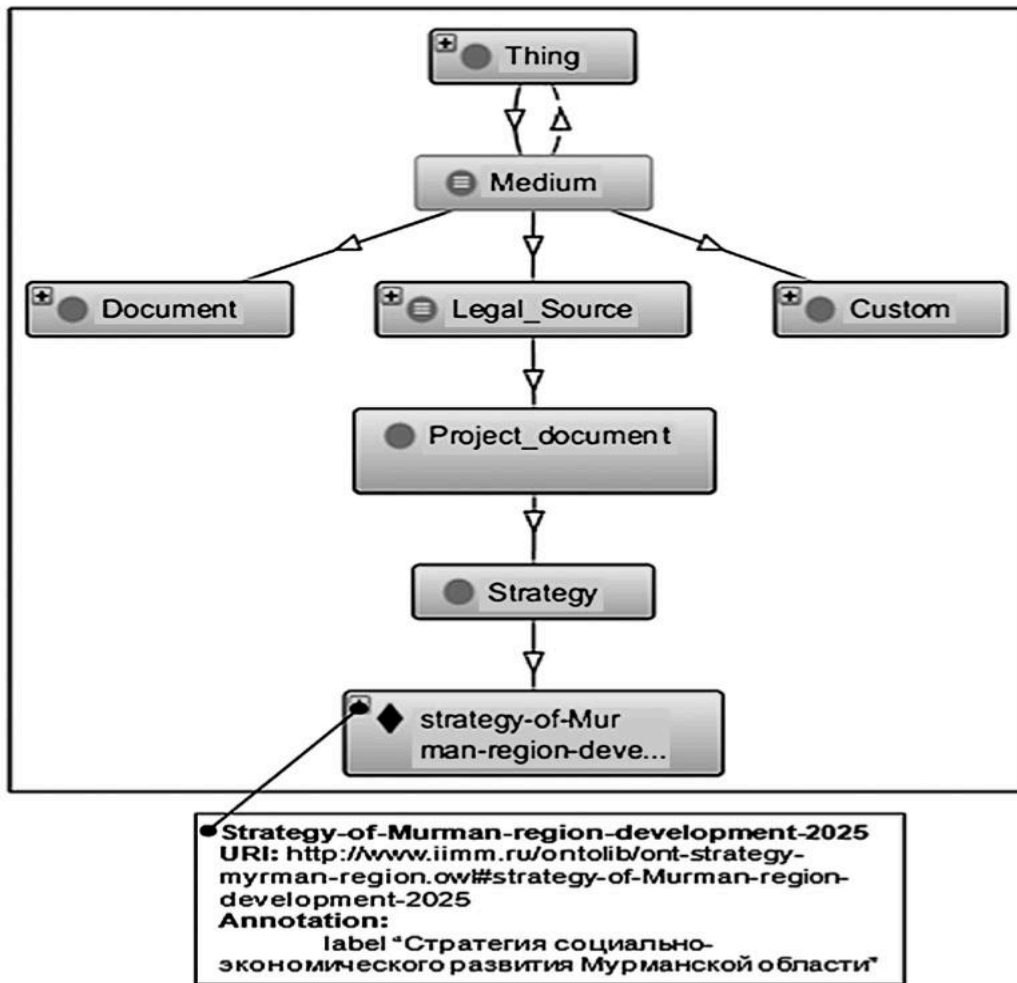
Основной идеей учёных была разработка автоматизированной системы проверки проектов нормативных актов на целостность и отсутствие противоречий.

В идеальном представлении авторов, такая система способствовала бы обеспечению целостности правового поля при принятии новых актов.

Решать данную задачу учёные планировали с помощью представления нормативных документов в OWL-онтологии, что обеспечило бы выявление некоторых отношений между классами, указывающих на неточности или противоречия в положениях исходных документов.

Методология формализации нормативно-правового документа в онтологии, по мнению учёных, должна была состоять из следующих основных шагов:

- определение документа как экземпляра класса «Основание» (Medium) или его подклассов, а также атрибутов документа: относительный уровень юридической силы, автор, дата публикации, дата вступления в силу, исполнитель и др.;
- представление в онтологии объектов, процессов, явлений реального мира, используемых в положениях документа;
- формализация положений документа в виде подклассов класса «Предложение» (Proposition) и/или суждений (Qualification);
- корректировка положений в соответствии с предложенными системой рекомендациями.



В ходе эксперимента П.А. Ломов и А.Г. Олейник подтвердили свою концепцию на отдельных примерах формализации некоторых положений документа и пришли к выводу, что онтология LKIF может быть принята за базовую версию концептуального описания юридической картины мира для поставленных целей.

Это лишь подтверждает наши выводы о том, что LKIF является крайне верхнеуровневой онтологией, которая подходит для формализации текстов законов на абстрактном уровне, но не применима для решения практических задач по обработке входных данных и формированию суждений и умозаключений с точки зрения юридической логики.

Онтология LKIF в области изучения моделирования правовых знаний является, пожалуй, самой исследуемой в научном сообществе. Ей посвящены следующие научные работы:

➤ М. Ceci, A. Gangemi («An OWL ontology library representing judicial interpretations», Semantic Web, 2016);

- Cevenini C., Contissa, G., Laukyte, M., Riveret, R., Rubino, R. («Development of the ALIS IP ontology: Merging legal and technical perspectives», Computeraided innovation (CAI), Boston, 2008);
- Distinto, I., d'Aquin, M., & Motta, E. («LOTED2: An ontology of European public procurement notices», Semantic Web, 2016);
- Ghosh, M. E., Naja, H., Abdulrab, H., & Khalil, M. («Towards a legal rule-based system grounded on the integration of criminal domain ontology and rules», In Procedia Computer Science: 112, 2017);
- Rodrigues, C. M. O., Azevedo, R. R., Freitas, F. L. G., da Silva, E. P., & da Silva Barros, P. V. (An ontological approach for simulating legal action in the brazilian penal code. In Proceedings of the 30th annual ACM symposium on applied computing», New York, 2015) и др.

Кроме того, онтология LKIF применялась в практико-ориентированных проектах. На ее основе предпринимались попытки систематизации источников права во Вьетнаме с возможностями поиска по базе нормативных актов, а также попытки структурирования правил пассажирских авиаперевозок на основе выборки наиболее нарушаемых прав пассажиров.

С 2008 года проект Estrella, в рамках которого разрабатывалась онтология LKIF, закрыт.

### 2.1.3. Legal Rule ML

Следующей моделью представления знаний, которую мы предлагаем рассмотреть, является Legal Rule ML. Данная модель разработана некоммерческим сообществом OASIS, осуществляющим реализацию проектов в области информационных технологий (кибербезопасность, блокчейн, защита информации, криптография, облачные вычисления, IoT и др.) с открытым исходным кодом с 1993 г. Само сообщество в своей деятельности стремится продвигать работу, направленную на снижение затрат и повышение эффективности разных видов деятельности, стимулирование инноваций, расширение рынков и развитие процессов мировой глобализации.

Одним из направлений деятельности сообщества стало развитие универсального языка представления юридических текстов. Интересен также тот факт, что над созданием Legal Rule ML работала группа учёных, ранее участвовавших в проекте Estrella по разработке LKIF Core ontology.

После закрытия проекта Estrella в 2008 г. и прекращения поддержки LKIF Core, Legal Rule ML по праву считается преемником LKIF, хотя и между ними очевидно не прослеживаются прямые признаки преемственности, кроме целей разработки.

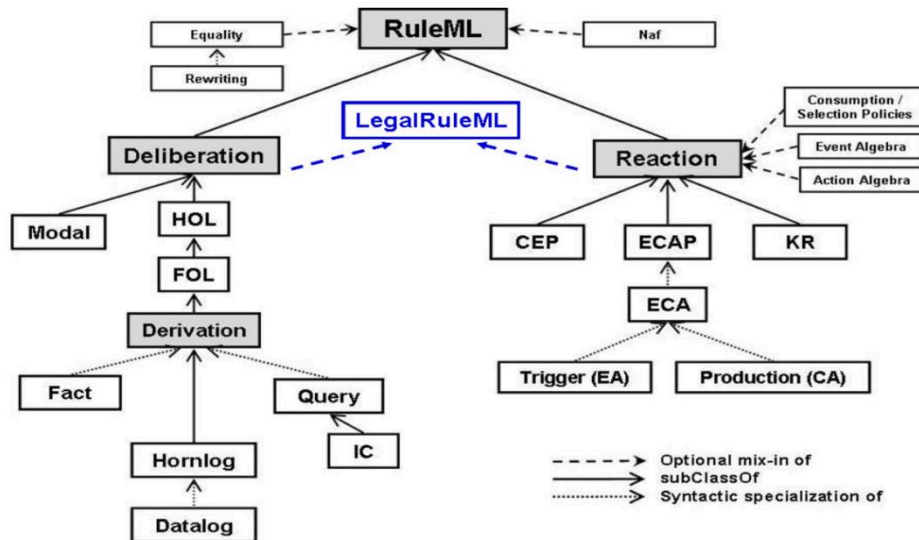
Как отмечают представители сообщества, на создание Legal Rule ML их побудило обнаружение ряда проблем в области нормотворчества. Юридические тексты (законы, положения, судебные акты, договоры и др.) содержат в себе принципы и правила поведения, которыми руководствуются субъекты различных общественных отношений.

При этом с течением времени и с усложнением общественных отношений количество нормативных актов растёт, что порождает для правоприменителей значительные сложности в поиске, толковании, сравнении их содержания и др.

Кроме того существует проблема и для лиц, разрабатывающих и принимающих новые нормативные акты, поскольку необходимость ручной проверки проекта документа на непротиворечие действующим источникам права не всегда является эффективной и не позволяет избежать правовых коллизий в будущем.

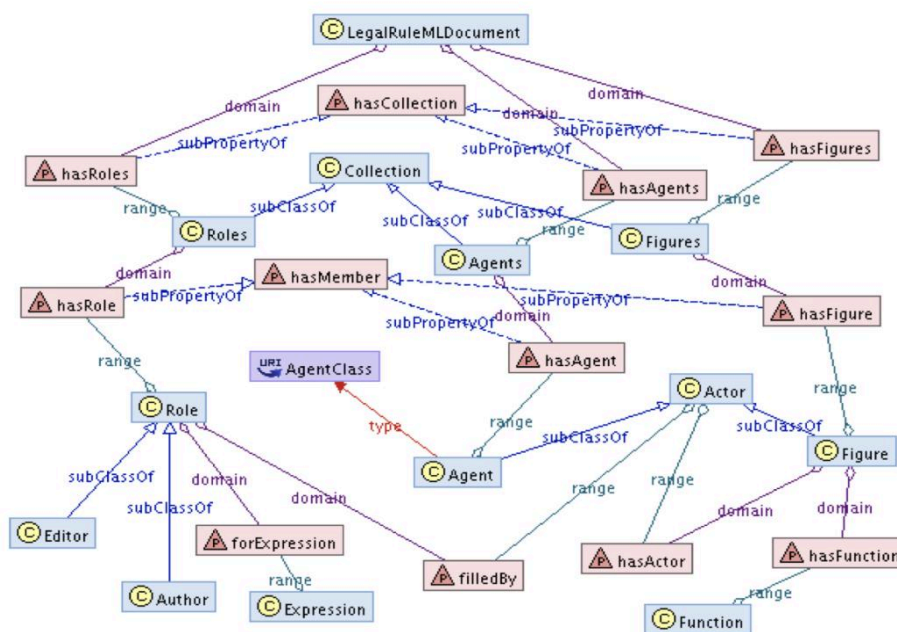
Потенциальным решением данных проблем, по мнению сообщества OASIS, может стать обеспечение возможности конвертации текстов нормативных актов в машиночитаемый формат (XML) и их структурирование. В последующем это станет платформой для унификации и структурирования юридических знаний, что будет способствовать развитию программных решений в области юриспруденции, а также позволит создать универсальный язык обмена правовыми знаниями и будет способствовать постепенному переходу от естественного языка при конструировании новых правовых норм к формализованному и машиночитаемому формату.

Именно для решения данных проблем была разработана Legal Rule ML.



Итак, с учётом поставленных создателями Legal Rule ML целей данная модель не является правовой онтологией в чистом виде, а представляет собой скорее язык обмена правилами на основе XML, который направлен на расширение более общего языка разметки правил Rule ML функциями, присущими юридической области. Legal Rule ML имеет верхнеуровневую структуру в виде групп Node elements и Edge elements, которые выполняют служебную роль по объединению групп правил.

Сами правила (значение нормы) выражаются с помощью существующих операторов (классов) в виде связей, отражающих воздействие нормы на субъекты или объекты.



В качестве основных операторов предусмотрены Agent (активный субъект, который действует или имеет возможность осуществлять действия), Figure (функция, которую выполняет субъект, или его статус) и Role (роль, которая формируется при реализации субъектом своих функций).

Например, с помощью данных операторов могут быть структурированы все нормативные акты, принятые конкретным должностным лицом.

Помимо данных категорий Legal Rule ML содержит операторы юрисдикции (географические пределы действия нормативного акта), полномочий, времени для отражения временных пределов действия нормативного акта (момент вступления в силу, момент прекращения действия и др.), а также событий как разновидности юридических фактов, с которыми могут быть связаны определённые правовые последствия.

Подводя итог обзору Legal Rule ML, отметим, что данная модель представления знаний, безусловно, имеет свои преимущества и позволяет структурировать содержание нормативного акта.

Но применение этой модели для целей правовой оценки данных невозможно в силу её специфической структуры. Мы не склонны считать это недостатком, поскольку изначальный замысел её создателей был направлен на выработку языка представления нормативного акта на основе логики и синтаксиса Rule ML.

Кроме того, Legal Rule ML по своей структуре не имеет привязки к правовой системе, поскольку оперирует общими понятиями и категориями, что позволяет использовать её для работы с нормативными документами разных государств.

Учитывая тот факт, что модель поддерживается и развивается сообществом OASIS в настоящее время, не исключено, что мировая общественность в будущем увидит впечатляющие практические результаты в данных направлениях, но не в контексте Legal AI и цифрового юриста.

#### **2.1.4. UFO**

Онтология Unified Fundamental Ontology (UFO) была разработана в 2004 году G. Guizzardi и его рабочей группой с целью разработки фундаментальной и универсальной теории концептуального моделирования любой области знаний. Данная модель не является примером юридической онтологии. Её ключевой задачей было создание некоторой обобщённой модели представления событий и действий объективной действительности, которую можно применять в самых разнообразных прикладных областях, в том числе в сфере права.

Первая версия UFO 0.1 появилась в результате синтеза двух других фундаментальных онтологий, а именно:

- General Formal Ontology (GFO), которая лежит в основе онтологического языка General Ontological Language (GOL) (разработка исследовательской группы OntoMed Research group при Университете Лейпциг (Германия));
- Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE), которая построена на основе принципов и методов OntoClean (разработка исследовательской группы ISTC-CNR-LOA Research group (Италия)).

Основу структуры UFO определяет систематический метод сравнения метамодели языка с конкретным представлением концептуализации той или иной предметной области, называемой эталонной онтологией.

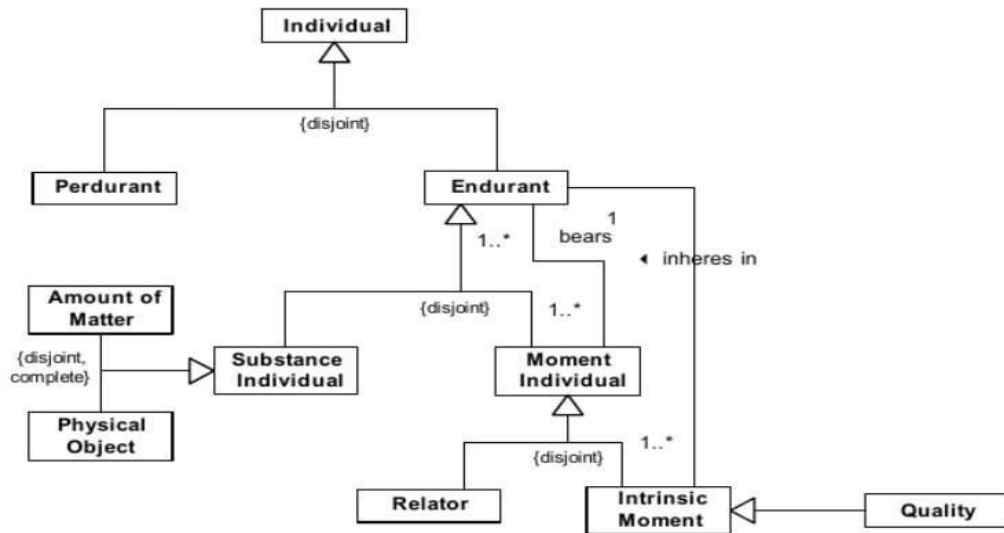
Онтология UFO успешно применяется для интерпретации моделей, включая бизнес-моделирование, а также для решения проблем семантической совместимости и интеграции различных языков моделирования. UFO является основополагающей онтологией для OntoUML, языка моделирования онтологий.

UFO состоит из трёх фрагментов: UFO-A, UFO-B, UFO-C. Основные категории UFO (UFO-A) были полностью формально охарактеризованы в докторской диссертации G. Guizzardi и далее расширены в исследовательской группе по онтологии и концептуальному моделированию (NEMO) в Бразилии совместно с сотрудниками Бранденбургского технологического университета (Gerd Wagner) и Лаборатории прикладной онтологии (LOA).

UFO-A был использован для анализа структурных концептуальных конструкций моделирования, таких как типы объектов и таксономические отношения, ассоциации и отношения между ассоциациями, роли, свойства, типы данных и зависимые сущности, а также парные отношения между объектами.

Более поздние разработки включают в себя онтологию событий в UFO (UFO-B), а также онтологию социальных и интенциональных аспектов (UFO-C).

Комбинация UFO-A, UFO-B и UFO-C использовалась для анализа, перепроектирования и интеграции эталонных концептуальных моделей в ряде сложных областей, таких как, например, моделирование предприятий, разработка программного обеспечения, нефтегазовая отрасль, телекоммуникации и биоинформатика.



В основу онтологии положено понятие фундаментального различия между двумя категориями: Particular (Individual) и Universal (Type). Particulars – сущности, которые существуют в реальности и обладают уникальной индивидуальностью. Universals, наоборот, представляют собой шаблоны характеристик, которые могут быть представлены несколькими различными индивидуальностями (Particulars).

В рамках фрагмента UFO-A также вводится понятие Substances – экзистенциально независимые индивидуальности (Particulars).

Понятие Момент (Moment), напротив, обозначает индивидуализированное свойство/характеристику, которое не может существовать без других индивидуальностей. Отношения (Relations) – это сущности, которые соединяют другие сущности. Relators - индивидуальности, обладающие возможностью соединять другие сущности.

Описываемые в части UFO-A индивидуальные сущности объединяются общим понятием – Endurants. Endurants – сущности, которые не имеют промежуточного состояния. Если они присутствуют, то присутствуют целиком и полностью.

В части UFO-B вводится еще одно обобщающее понятие – Perdurants (события), представляющее собой сущности, состоящие из временных частей. Они «случаются во времени» в том смысле, что они распространяются в течение времени, накапливая временные части.

В отличие от Endurants, если Perdurant присутствует, то это ещё не означает, что все его составляющие временные части также присутствуют. Основную категорию в онтологии UFO-B занимает концепт «Событие» (Event), он же Perdurant. События могут изменять реальность: в результате возникновения события

реальность переходит от одной ситуации (предшествующей наступлению события) к другой (после наступления события).

С одной стороны, события экзистенциально зависят от непосредственных участников данных событий, с другой стороны, события можно рассматривать с точки зрения временного аспекта и расширяющих его сущностей.

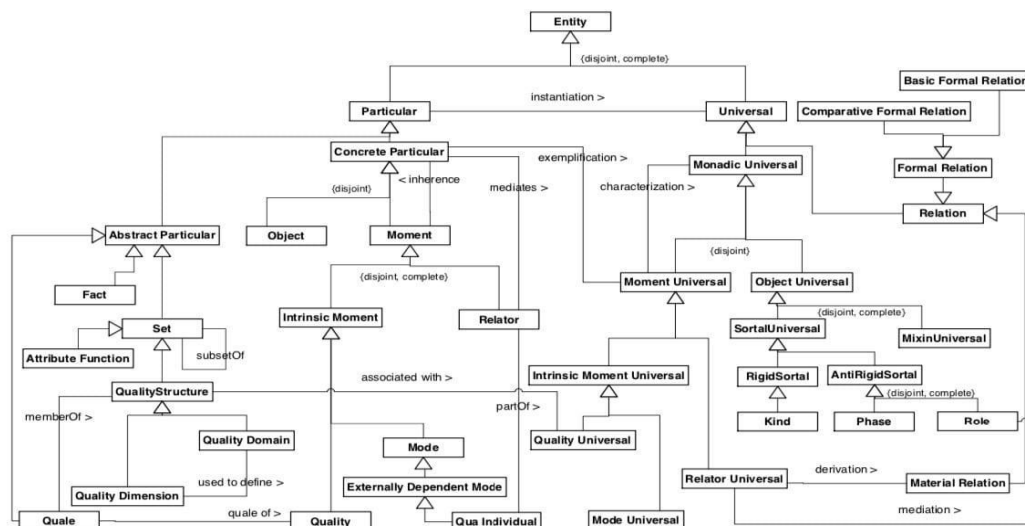
На третьем уровне онтологии UFO-C рассматриваются социальные сущности (Endurants и Perdurants). Вводятся понятия Agents (агенты) и Objects (объекты), которые представляют собой экзистенциально независимые индивидуальности, в первом случае – деятельные (агентивные), во втором случае – бездействующие, как правило, неодушевлённые объекты. Агенты могут «нести» так называемые преднамеренные моменты (Intentional Moments). Каждый Intentional Moment имеет тип (например, Belief, Desire, Intention) и содержание (Proposition). Намерение – это тип интенциональности, называемый Intention. Содержание намерения (propositional content) – это цель (Goal).

Два других типа интенциональности – намерение (Belief) и желание (Desire). Желание выражает стремление агента к определённом состоянию в реальности, тогда как намерение – желаемое состояние, для достижения которого агент готов выполнить какие-то действия.

По этой причине намерения вынуждают агентов выполнять действия (Actions).

Действия – это события, которые являются Instance некоторого плана (Plan), и имеют своей целью удовлетворение некоторого намерения.

Примером атомарного действия является акт коммуникации (Communication Act), который может быть использовать для создания социальных моментов (Social Moments). Комплексное действие, для выполнения которого необходим вклад нескольких агентов, называется взаимодействием (Interaction).



В целом, если оценивать концепцию представления объективной реальности в онтологии UFO, то можно отметить ее универсальность и преемственность. Многие юридические концепты могут быть описаны с помощью такой логики.

Примечательно, что в научном сообществе еще в 2015 г. предлагались идеи адаптации UFO под правовую реальность путем разработки нового блока UFO-L, который бы детализировал существующую онтологию конкретными правовыми категориями.

К сожалению, нам не удалось найти публикации о результатах таких экспериментов, однако перспективы успеха в данных начинаниях имеют место быть. UFO была объектом исследования в таких научных работах, как:

- Griffo, C., Almeida, J. P. A., & Guizzardi, G. A systematic mapping of the literature on legal core ontologies. *Ontobras*. In *CEUR Workshop Proceedings*: 1442. CEUR-WS.org, 2015;
- Rodrigues, C. M. O., Freitas, F. L. G., & Azevedo, R. R. An ontology for property crime based on events from ufo-b foundational ontology. *5th Brazilian conference on intelligent systems (bracis)*, 2016 и др.

### 2.1.5. FBO

Последней онтологией в данном обзоре является Frame-based ontologies (FBO), которая представляет собой совокупность онтологий, включающую в себя общую юридическую онтологию (нормы, акты, описания концептов) и онтологию, которая приобретает статус так называемой специфической онтологии. Различие основано на наблюдении, что некоторые части онтологии могут быть повторно использованы в различных правовых поддоменах.

Общая правовая онтология (GLO), в отличие от специфической онтологии, является универсальной и многоразовой частью онтологии.

Она делит юридические знания на три отдельных объекта: нормы, акты и описания концептов. Для каждого из этих объектов онтология определяет шаблон (также называемый структурой кадра), в котором перечислены все атрибуты, относящиеся к объекту.

Нормы являются общими правилами, стандартами и принципами поведения, которые субъекты права обязаны соблюдать.

В онтологии норма включает в себя следующие восемь элементов:

- идентификатор нормы (используется в качестве отправной точки для нормы);
- тип нормы (либо норма поведения, либо норма компетенции);
- обнародование (источник нормы);
- сфера охвата (сфера применения нормы);
- условия применения (обстоятельства, при которых применяется норма);
- субъект нормы (физическое или юридическое лицо, которому адресована норма);
- юридическая форма (должна, не должна, может, не может);
- идентификатор действия (используется в качестве ссылки на отдельное описание действия).

Акты представляют собой динамические аспекты, которые влияют на изменения в состоянии мира.

В рамках категории актов существуют два различия.

Первое различие между событиями и процессами. События представляют собой мгновенное изменение между двумя состояниями, в то время как процессы имеют продолжительность.

Второе различие заключается между институциональными актами и физическими актами. Оно выражается в том, что первый тип актов рассматривается как правовые (институциональные) версии (физических) актов, которые происходят в реальном мире (институциональный акт – это юридическая квалификация физического акта). Предполагается, что все акты имеют следующие тринадцать элементов:

- идентификатор акта (используется в качестве отправной точки для акта);
- обнародование (источник описания акта);
- область действия (область применения описания акта);
- агент (физическое лицо, группа лиц, совокупность или конгломерат);
- тип акта (как базовые акты, так и действия, которые были указаны в другом месте);
- модальность средств (материальные объекты, используемые в акте);
- модальность манеры (способ, которым использовались объекты или выполнялись действия);

- временные аспекты (спецификация абсолютного времени, например, на первое августа, по воскресеньям, ночью и т. д., но не во время пожара, после смерти президента и т. д.);
- пространственные аспекты (указание места, где происходит акт);
- косвенные аспекты (описание обстоятельств, при которых происходит действие);
- причина действия (указание причины выполнить действие);
- цель действия (цель, визуализируемая агентом);
- намеренность действия (состояние ума агента);
- конечное состояние (результаты и последствия действия).

Концептуальные описания имеют дело со значениями понятий.

Они могут быть определениями или определяющими положениями и могут использоваться для окончательного определения значения понятия, либо путем обеспечения необходимых и достаточных условий.

Другим типом концепции является фактор, который может либо установить достаточное условие, либо указать некоторый вклад в применимость концепции.

Наконец, существуют мета-концепции, которые являются положениями, регулирующими применение других положений.

Специфическая онтология состоит из предикативных отношений, которые используются в качестве дополнения к терминологии для норм, актов и описаний понятий.

Специфическая онтология не может быть повторно использована для других юридических поддоменов, и её всегда следует создавать для каждого рассматриваемого юридического поддомена. Специфическая онтология устанавливает словарный запас, с помощью которого создаётся база знаний. Ключевое различие между GLO и специфической онтологией состоит в следующем:

- различие между физическими актами и индивидуальными актами (юридическое – интерпретация физических актов, которые происходят в реальном мире);
- акцент на модальность действия, временные и пространственные аспекты и обстоятельства;
- представление объёма, условий применения и правовой формы норм.

При оценке практической применимости онтологии FBO для целей юридической экспертизы мы приходим к выводам, что данная модель нацелена на структурирование и формализацию текстов нормативных актов с помощью GLO.

Однако наличие специфической онтологии в ее составе позволяет выражать влияние норм права на субъекты и объекты материального мира, а именно какие действия субъектов и какое воздействие на объекты опосредованы и связаны с определённой нормой, содержащейся в тексте нормативного акта.

Данное обстоятельство, безусловно, расширяет общую применимость FBO, но не для целей Legal AI, поскольку в последнем случае требуется глубокая детализация правовых концептов, а не правовых норм.

## **2.2. Результаты сравнения**

В рамках обзора мы задались целью проанализировать и оценить юридические онтологии, которые создавались в разные периоды времени.

В результате мы пришли к выводу, что общим для большинства данных моделей представления юридических знаний является их привязанность к описанию законов и норм.

Они не описывают конкретные нормативные документы, а содержат универсальные концепты их представления. Многие из онтологий изначально создавались для концептуализации самой структуры нормативно-правовых актов и представления их содержания на языке семантической структуры.

Обзор представленных онтологий привел нас к выводу о том, что такие онтологии выполняют задачу обмена знаниями, поиска противоречий и дублирования норм, но никак не служат основой для формализации правовых знаний для систем юридического искусственного интеллекта.

Результаты сравнения и оценки правовых онтологий представлены в таблице ниже.

Оценка производится на основе 5-бальной шкалы по 6 критериям.

### **1. Уровень детализации.**

Данный критерий отражает степень детализации правовых категорий и терминов, с помощью которых в онтологии могут быть структурированы правовые знания.

На данный критерий влияет количество классов в онтологии, а также количество и разнообразие связей.

При этом значение имеют те классы и связи, которые позволяют выражать именно правовые знания, а не общие концепты объективной реальности.

Оценка по данному критерию производится следующим образом:

- 0 – полное отсутствие необходимых классов и связей;
- от 1 до 2 – наличие общих классов и связей, позволяющих на абстрактном уровне отражать базовые юридические категории (например, организация, место и др.);
- от 3 до 4 – наличие системы специальных правовых классов и связей, позволяющих отражать юридические концепты;
- 5 – наличие развитой системы правовых классов и связей, позволяющих отражать юридическую картину мира в полной мере.

## **2. Возможность структурирования концептов отечественного права в текущей версии.**

Данный критерий отражает возможность представления правоотношений в существующей в системе отечественного права при использовании исходной версии онтологии.

Иными словами он отражает степень пригодности онтологии для права РФ.

Оценка по данному критерию производится следующим образом:

- 0 – полная невозможность использования;
- от 1 до 2 – наличие общих классов и связей, которые могут быть использованы при определённых ограничениях и допущениях;
- от 3 до 4 – наличие системы специальных правовых классов и связей, позволяющих отражать некоторые отечественные концепты;
- 5 – полная совместимость с отечественным правом.

### **3. Возможность доработки под особенности отечественного права.**

Данный критерий отражает степень привязанности структуры онтологии к определённой правовой системе и возможность её модификации под особенности и характеристики права РФ.

Оценка по данному критерию производится следующим образом:

- 0 – полное отсутствие возможности адаптации;
- от 1 до 2 – слабо выраженная возможность адаптации при сохранении существенных ограничений и допущений;
- от 3 до 4 – наличие возможности адаптации, которая требует значительной переработки структуры;
- 5 – возможность простой адаптации при незначительной доработке.

### **4. Логика и удобство использования.**

Данный критерий отражает степень логичности существующей структуры онтологии и сложность ее восприятия и использования специалистом предметной области.

Оценка по данному критерию производится следующим образом:

- 0 – отсутствие логики и сложность для восприятия;
- от 1 до 2 – наличие определённой логики построения онтологии, но сложность ее восприятия;
- от 3 до 4 – наличие выраженной и понятной логики, которая доступна специалисту предметной области при изучении онтологии;
- 5 – логичность структуры и простота восприятия.

### **5. Ширина охвата правоотношений.**

Данный критерий отражает количество правоотношений, которое может быть структурировано в онтологии при существующей степени её детализации.

В отличие от первого критерия данная характеристика отвечает на вопрос: «Насколько онтология приближена к разнообразию реальных правоотношений?».

Оценка по данному критерию производится следующим образом:

- 0 – полное отсутствие концептов для структурирования правоотношений;
- от 1 до 2 – наличие общих классов и связей, позволяющих на абстрактном уровне отражать базовые правоотношения;
- от 3 до 4 – наличие системы специальных правовых классов и связей, позволяющих отражать некоторые сложные правоотношения;
- 5 – наличие развитой системы правовых классов и связей, позволяющих отражать все разнообразие правоотношений.

#### **6. Зависимость от изменений законодательства.**

Данный критерий отражает степень привязанности онтологии к действующему в настоящий момент или на момент ее создания законодательству.

Необходимость введения данного критерия обусловлена тем, что правовая система РФ, в отличие от многих зарубежных правовых систем, находится на стадии своего формирования и развития, в связи с чем достаточно часто принимаемые нормативные акты коренным образом изменяют и/или дополняют регулирование определенной группы отношений. Оценка по данному критерию производится следующим образом:

- 0 – строгая зависимость и необходимость полной переработки при изменении законодательства;
- от 1 до 2 – наличие сильной зависимости, которая требует различных изменений структуры онтологии при изменении законодательства;
- от 3 до 4 – наличие средней зависимости и необходимость изменения только при фундаментальных реформах;
- 5 – слабая зависимость и отсутствие необходимости модификации.

№	Критерий	FIBO	LKIF	Legal Rule ML	UFO	FBO
1.	Уровень детализации	3	2	2	1	2
2.	Возможность структурирования концептов отечественного права в текущей версии	2	1	1	1	1
3.	Возможность доработки под особенности отечественного права	3	1	1	1	1
4.	Логика и удобство использования	5	2	2	3	2
5.	Ширина охвата правоотношений	3	0	0	0	0
6.	Зависимость от изменений законодательства	4	5	5	5	5
<b>Итого</b>		20	11	11	10	11

Онтологии LKIF, Legal Rule ML, FBO изначально создавались для представления текстов нормативных актов, поэтому по своей структуре они адаптированы под формализацию законов.

Обладая высокой степенью абстракции, данные модели потенциально позволяют описать практически любой нормативный акт во многих его частях независимо от правовой системы.

При этом верхнеуровневость онтологий не привязывает их к действующему законодательству, правовой системе и др., придавая им свойства универсальности и «reusable».

Однако отсутствие возможности описать реальные правовые концепты, роль участников правоотношений, взаимосвязи и свойства субъектов, объектов права и др. не позволяет использовать их для Legal AI.

Исключением служат онтологии UFO и FIBO. UFO является универсальной моделью отражения отношений в любой области знаний, что потенциально позволяет использовать её для юридической сферы.

Однако к настоящему времени модуль UFO-L не разработан, поэтому оценивать эффективность UFO в области права невозможно.

FIBO в значительной степени является практико-ориентированной финансовой онтологией, и ее эффективность в решении практических задач многократно подтверждена.

Однако предметная область данной онтологии нацелена на отражение взаимоотношений в области финансов и бизнеса, что не позволяет применять её к юриспруденции, поскольку данные сферы хотя и пересекаются, но имеют различное и самостоятельное содержание. Кроме того, уровень детализации в FIBO невысокий, что также ограничивает сферу её применения.

Таким образом, мы приходим к выводу, что существующие онтологии в области права не подходят для решения задач Legal AI в силу низкого уровня детализации, нацеленности не на систематизацию правовых знаний, а на структурирование нормативных актов, уникальности правовых систем и невозможности создания универсальной юридической онтологии для применения во многих странах.

Данный вывод свидетельствует о необходимости самостоятельного создания юридической онтологии (графа знаний), который бы соответствовал практическим потребностям Legal AI и правовому полю РФ.

И здесь напрашивается очевидный вопрос: «Какие подходы и критерии следует использовать для создания необходимой базы юридических знаний?».

В качестве примера онтологии, которая, по нашему мнению, является эталонной моделью структурирования знаний, хотелось бы привести онтологию объектов культурного наследия CIDOC CRM (Committee on Documentation «Conceptual Reference Model»).

CIDOC CRM разработана рабочей группой по стандартизации документации комитета CIDOC и специализированной рабочей группой.

Она представляет собой формальную онтологию, предназначенную для улучшения интеграции и обмена гетерогенной информацией по культурному наследию и является онтологией верхнего уровня для данной предметной области.

Если рассматривать более конкретно, CIDOC CRM определяет семантику схем баз данных и структур документов, используемых в культурном наследии и музейной документации, в терминах формальной онтологии.

Модель не определяет терминологию, используемую в конкретных структурах данных, но имеет характерные отношения для её использования. Она не стремится предлагать то, что должны документировать учреждения культуры, но объясняет логику того, что они фактически документируют и представляют семантическую интероперабельность между музеями, библиотеками, архивами.

В настоящий момент онтология используется в ряде крупных проектов, таких как MIDAS XML, IST Project SCULPTEUR, IST Project I-Mass и др.

Модель может служить как руководством для разработчиков информационных систем, так и общим языком для экспертов предметной области и специалистов по информационным технологиям, чтобы сформулировать требования к информационным системам и служить руководством для надлежащей практики концептуального моделирования.

Она предназначена для покрытия контекстной информации исторического, географического и теоретического характера об отдельных экспонатах и музейных коллекциях в целом.

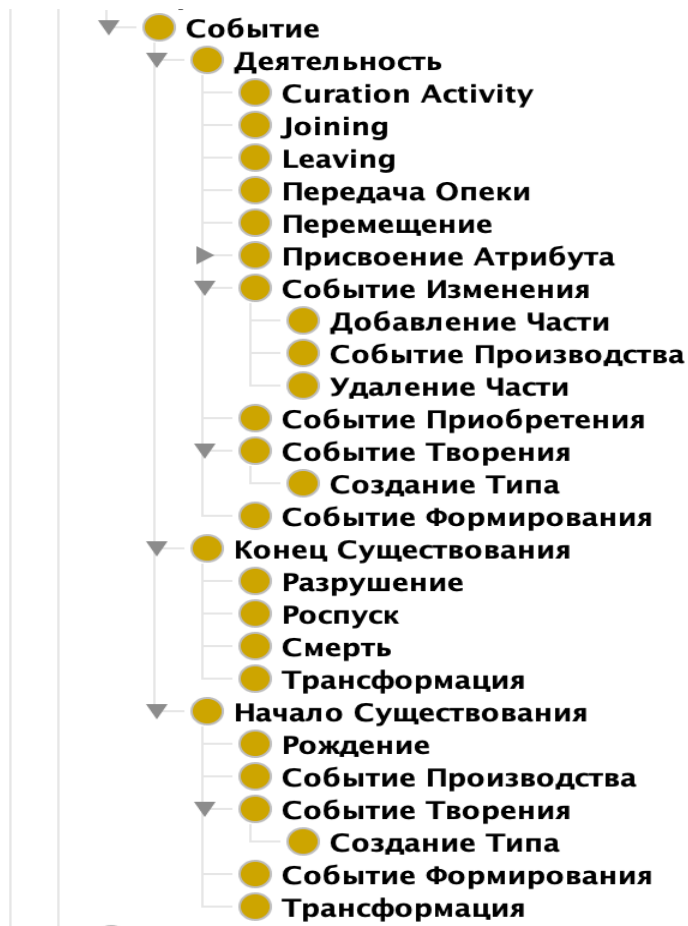
Структурно CIDOC CRM состоит из иерархии классов и широкого набора свойств (бинарных отношений), связывающих классы между собой. Версия 5.0.1 CIDOC CRM содержит 82 класса и 142 свойства, связывающих классы между собой и описывающих предметы, понятия, людей, события, место, время и их отношения.

Все концепты (классы и свойства) модели можно разделить на три группы.





Первая группа включает классы и отношения, охватывающие наиболее общие понятия окружающего мира: постоянные и временные сущности, отношения участия, зависимости, совпадения во времени.



Вторая группа содержит понятия, частично поддерживающие функции управления: приобретение и учёт единиц хранения, передача прав собственности на объекты культуры.

К третьей группе относятся классы и свойства, используемые для внутренней организации самой онтологии: средства, необходимые для подключения внешних источников терминов, например, тезаурусов по отраслям культуры.

Иерархия классов модели CIDOC CRM делится на 2 ветви: Постоянные сущности и Временные сущности.

Прочие классы являются вспомогательными. На самых нижних уровнях иерархии классов появляются понятия, характерные для сферы культуры: «Хранение», «Перемещение (ценностей)», «Проект или Процедура» (в том числе техника производства), «Период» (в том числе художественный стиль).

Иерархия классов может быть гибко расширена с применением встроенного класса «Тип».

Наибольший интерес представляют свойства.

Классы на нижних уровнях иерархии имеют около 10-15 свойств, причём большая часть свойств наследуется от классов-родителей.

Названия свойств представляют собой глагольные фразы, выбранные так, что при последовательном связывании двух классов свойством получается осмысленная фраза с субъектом, предикатом и объектом.

Отличительным свойством данной онтологии является то, что она не преследует цель описать документы музейных экспонатов или сами объекты культурного наследия.

Онтология воспроизводит свойства и отношения таких объектов, позволяя описать всю полноту характеристик любых музейных экспонатов в универсальном и формализованном формате.

Именно такими свойствами, по нашему мнению, должна обладать юридическая онтология.

- ▶  появился в присутствии
- появляется во течение
- появляется до
- появляется после
- ▶  право принадлежит
- предоставляет доступ к
- предшествует во времени
- ▶  признаки также найдены на
- применяется к
- принадлежит
- ▶  присвоил
- ▶  присвоил атрибут для
- ▶  присутствовал при
- равен по времени
- сделан из
- следует во времени за
- сменил владельца через
- ▶  совершался на
- содержит
- содержит
- содержит
- содержится в
- содержится в
- составлен из
- составлен из
- ▶  составлен из
- состоит из
- состоит из
- ▶  ссылается на
- формирует часть
- ▶  формирует часть
- формирует часть
- формирует часть
- формирует часть
- ▶  является бывшим или текущим владельцем для
- ▶  является бывшим или текущим зрителем для
- является величиной для
- является временным отрезком для
- является действующим или бывшим членом
- является единицей для
- является объектом для
- является текущим или бывшим местом жительства для
- ▶  является текущим или бывшим местоположением для
- является текущим постоянным местоположением для
- ▶  является типом для
- является условием для
- является языком для
- являлся местом для

- был задуман для
- был интенцией для
- был максимальной длительностью для
- был местом совершения
- был минимальной длительностью для
- был отцом для
- был предполагаемым использованием для
- был применением для
- был присвоен посредством
- был создан для
- был типом объекта использованного в
- был целью для
- был целью для
- включает
- владеет
- владеет правом на
- входит в состав
- граничит с
- демонстрирует признаки
- заканчивает
- заканчивается
- идентифицирует
- идентифицируется посредством
- имеет альтернативную форму
- имеет бывшего или текущего владельца
- имеет бывшего или текущего зрителя
- имеет величину
- имеет временной отрезок
- имеет вышестоящий термин
- имеет действующего или бывшего члена
- имеет длительность меньше чем
- имеет длительность по крайней мере больше чем
- имеет единицу
- имеет контакт
- имеет определение района
- имеет район
- имеет ссылку на себя от
- имеет текущее или бывшее местожительства
- имеет текущее или бывшее местоположение
- имеет текущее постоянное местоположение
- имеет тип
- имеет условие
- имеет язык
- имел конкретную цель
- имел место на или в
- имел общую цель
- имел основное применение
- использовал

Отдельно хотелось бы отметить необходимость использования *Ontology Design Patterns*. Моделированием онтологий научное сообщество занимается уже не один десяток лет, и за данный продолжительный период времени были разработаны многочисленные концепты, которые нашли свое успешное применение в различных моделях знаний и сферах.

Дело в том, что несмотря на различия предметных областей (право, финансы, медицины, культура и др.), внутри них существует много общего.

Концептуальные отношения между объектами (часть и целое, принадлежность, зависимость и др.) мы встречаем в любой области знаний.

Кроме того, даже в каждой узкоспециализированной онтологии нельзя обойтись без моделирования общих категорий, позволяющих отражать те или иные состояния действительности. Речь идёт, например, о времени, событиях и др.

Так, существует некоммерческий проект *ODPA (Ontology Design Patterns)*, который предлагает использовать выработанные и универсальные концепты записи отдельных объектов и категорий в онтологическом формате, которые позволяют улучшить методологию и повысить эффективность создаваемой онтологии.

## **6. Автогенерация онтологий**

В предыдущей публикации мы упоминали в качестве альтернативной методологии построения моделей знаний автогенерацию онтологий. Автоматическая генерация онтологий представляет собой процесс формирования онтологии предметной области на основе обучающей выборки документов в автоматическом режиме с последующей ручной корректировкой отдельных понятий и связей.

Основным аргументом сторонников данного подхода является то, что моделирование юридической онтологии в ручном режиме потребует многолетней работы, а существующий инструментарий по извлечению данных из текста и их пост-обработке позволит сгенерировать части необходимой онтологии с минимумом временных затрат.

Мы выступаем против данного подхода, поскольку для построения юридической онтологии и графов автогенерация не применима, и тому есть несколько подтверждений.

Во-первых, ранее мы говорили о том, что ключевое значение в юридической онтологии имеют связи, поскольку именно они отражают множественные отношения между классами. Онтологии и графы, создаваемые для *Legal AI*, должны

иметь в совокупности сотни и тысячи связей для решения даже простых юридических задач.

Если воспользоваться математической формулой полного графа  $n*(n-1)/2$ , где  $n$  – количество задействованных классов, вершин, мы приходим к выводу, что необходимое количество связей на порядок превышает количество необходимых классов.

При автогенерации онтологии на выходе мы получаем крайне скудный набор связей, которого явно недостаточно для полноценного описания юридической картины мира.

В частности, среди всего многообразия отношений при автогенерации онтологии мы получаем только отношения принадлежности (is-a, subclass of).

Во-вторых, существует проблема в источнике данных для генерируемой онтологии: не существует такого набора документов, который бы наиболее полно и целостно описывал непосредственную юридическую жизнь (legal practice). Использование текстов нормативных актов и судебных решений – метод, который, на наш взгляд, заведомо обречён на неудачу, поскольку, как мы уже упоминали ранее, законы и судебная практика несут в себе не более 30% всех правовых знаний.

Кроме того, цена ошибки в данном случае слишком высока. Даже тексты существующих судебных решений в полной мере не отражают смысловое содержание нормативных актов и реальную юридическую практику.

Например, процедуре принятия искового заявления арбитражным судом посвящено 6 статей АПК РФ (ст. 125-129 АПК РФ).

Однако при построении графа, отражающего данную процедуру и отвечающего только на два вопроса (может ли данное исковое заявление быть принято к производству арбитражным судом, и какой судебный состав уполномочен рассматривать данное исковое заявление), данных статей было недостаточно для выполнения поставленной задачи. На множество вопросов тексты нормативных актов (не только АПК РФ, но и судебные регламенты, судебная практика и др.) не дают ответов, поэтому необходимо руководствоваться опытом практикующих юристов, который документально нигде не оформлен.

В-третьих, если обратиться к предыдущим рассуждениям о иерархической структуре знаний (пирамида Рассела Акоффа), в которой выделяются data (неструктурированные, разрозненные данные), information (структурированные данные, информация), knowledge (знания) и wisdom (мудрость), то становится очевидным тот факт, что инструменты ML и DL работают на 90 % с неструктурированными данными и только на 10% с информацией.

Однако ценные знания и опыт предметной области находятся на уровнях *knowledge* и *wisdom*, которые существуют только в сознании экспертов соответствующей области знаний. Суммарно данные факторы объясняют отсутствие успешных попыток автоматической генерации онтологий.

Также важно понимать, что концепция машинного обучения исходит из попыток создать инструмент, который получает исходные данные, а на выходе выдаёт правильный результат. В случае с автогенерацией правовой онтологий применение машинного обучения – это попытка создать *end-to-end* решение, которое на основе выборки текста (законов, судебных решений и др.) смоделирует необходимую концептуальную модель знаний.

Однако в силу высокой сложности данной задачи и необходимости задействовать *common sense* такой подход при нынешнем уровне развития технологий представляется тупиковым.

## **7. Заключительные положения**

В результате обзора существующих онтологий мы приходим к главному выводу: если для зарубежных правовых систем существуют различные варианты готовых решений, которые в разной степени соответствуют потребностям *Legal AI*, то для отечественной системы права моделей юридических знаний не существует вовсе.

Причин тому множество, одной из них является отсутствие значительных государственных инвестиций в развитие данного направления. А ведь создание структурированной юридической картины мира в формате графа знаний позволит решить наиболее сложную задачу при разработке *Legal AI* – обеспечить искусственный интеллект логикой и знаниями эксперта соответствующей предметной области. Без решения данной задачи невозможно создать программный инструмент, способный генерировать суждения и умозаключения по итогам анализа определенных входных данных (*reasoning*). Именно поэтому следующим важным шагом после реализации инструментов лингвистической обработки текста является построение графа знаний юридической области, основанного на отражении объективной юридической реальности.

Сегодня в проектах по развитию систем искусственного интеллекта основной фокус сосредоточен на *ML* и *DL*.

Однако, как мы отмечали ранее, все машинное обучение работает на уровне неструктурированных данных (*data*) – уровне *Association*, что не позволяет перейти к решению интеллектуальных и сложных практических задач, связанных с

полноценной цифровой экспертизой. Для создания настоящего цифрового эксперта требуется переход на следующие уровни – knowledge (знания) и wisdom (мудрость). Перейти на данные уровни позволит, по нашему мнению, создание графа знаний юридической области, который станет основой Legal AI.

Помимо этого, важно понимать взаимосвязь между инструментами и технологиями, применяемыми для создания Legal AI. Само по себе создание онтологий и графов знаний не имеет практического смысла без разработки инструментария для процессинга текста на естественном языке.

Модели представления знаний служат источником логики предметной области и позволяют генерировать суждения и умозаключения на ее основе.

Однако входными данными для них являются результаты лингвистической обработки текста. Чтобы запустить reasoning, необходимо задать необходимые instances в онтологии, определить их исходные свойства, атрибуты, что без семантического анализа текста на естественном языке сделать невозможно.