

Я. Менекер\*

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ОТ ЭТИКИ К ПРАВОВОМУ РЕГУЛИРОВАНИЮ, ОСНОВАННОМУ НА ПРИНЦИПЕ ЗАЩИТЫ ПРАВ ЧЕЛОВЕКА<sup>1</sup>

**Аннотация.** Влияние технологий искусственного интеллекта на сегодняшний день является значительным. Искусственный интеллект — не просто новый артефакт, новые технологии уникальны тем, как они преобразуют наше отношение к информации, межличностным связям, как они трансформируют общество в целом. Такое системное влияние новейших технологий на многие аспекты общественной жизни оправдывает необходимость изучения пределов правового регулирования искусственного интеллекта в целях предотвращения рисков, связанных с его использованием.

Данное исследование посвящено рассмотрению актов так называемого мягкого права (не имеющего обязательной юридической силы), в частности документов Совета Европы, и ставит своей целью обоснование разработки соответствующей нормативно-правовой базы, обеспечивающей сбалансированность рисков, связанных с использованием технологий искусственного интеллекта. В рамках исследования также затронут вопрос о введении в действие соответствующих правовых инструментов, способствующих тому, чтобы нормативное регулирование было ясным, понятным и эффективным.

**Ключевые слова:** алгоритмы, машинное обучение, детальное исследование, САНАИ, сертификация, специальный комитет по искусственному интеллекту, Совет Европы, конвенция, рамочная конвенция, Конвенция 108, демократия, права человека, верховенство права, этика, этика искусственного интеллекта, искусственный интеллект, машинное обучение, регулирование, риски, стандарты.

**Для цитирования:** Менекер Я. Искусственный интеллект: от этики к правовому регулированию, основанному на принципе защиты прав человека // Право и цифровая экономика. — 2021. — № 2 (12). — С. 39—44. — DOI: 10.17803/2618-8198.2021.12.2.039-044.

Появление технологий искусственного интеллекта (далее — ИИ)<sup>2</sup>, пожалуй, стало одной из главных проблем текущего столетия. Знакомясь с исследовательскими работами<sup>3</sup> времен зарождения кибернетики, в которых уже звучало понятие ИИ, можем прийти к

© Менекер Я., 2021

\* Менекер Янник, французский магистрат при Совете Европы, советник по вопросам цифровой трансформации и искусственного интеллекта, научный сотрудник Института перспективных исследований в области правосудия (IHEJ — Institut des Hautes Etudes sur la Justice), Франция  
lab.kpuryr@msal.ru

<sup>1</sup> Статья подготовлена на основе доклада, сделанного в рамках Международной конференции «Юридическая конвергенция в условиях цифровизации», прошедшей 25 ноября 2020 г. в Московском государственном юридическом университете имени О.Е. Кутафина (МГЮА).

Высказанные в статье мнения представляют собой точку зрения исследователя и не являются официальной позицией Совета Европы.

<sup>2</sup> Все технологии, охватываемые термином «искусственный интеллект» (ИИ), не относятся к автономной личности, и во избежание любого антропоморфизма данный термин был выбран для обобщения более подходящих понятий: «инструменты искусственного интеллекта» или «приложения искусственного интеллекта».

<sup>3</sup> Sadin E. L'intelligence artificielle ou l'enjeu du siècle, anatomie d'un antihumanisme radical, L'échappée, 2018.

выводу, что данный термин вновь вошел в научный оборот с 2010 г.

Именно с этого времени ИИ стал постепенно проникать во многие сферы человеческой жизни. В период «пандемийного» кризиса 2020 г. технологии ИИ активно использовались для разработки тестов на выявление вируса COVID-19, поиска вакцины или выявления заболевания с помощью автоматического считывания показаний снимков легких<sup>4</sup>.

Вместе с тем необходимо отметить, что общественность выражает серьезную обеспокоенность по поводу развития систем ИИ с функцией автономного принятия решений, массовой слежки, приводящей к дискриминационным практикам.

Так, до недавнего времени использование алгоритмов на базе ИИ для автоматического генерирования результатов школьных экзаменов рассматривалось в качестве дискриминационной практики, поскольку способствовало усилению неравенства между школьниками<sup>5</sup>.

ИИ может ошибиться и принять голый череп футбольного судьи за футбольный мяч, что вызовет, скорее, улыбку на лице<sup>6</sup>, но такие действия в то же время демонстрируют ограниченные возможности новейших технологий и высокую вероятность возникновения ошибок, например при идентификации (распознавании) лиц.

Тем не менее необходимо уточнить: использование технологий ИИ не является универсальным решением всех проблем. Трудности, с которыми приходится сталкиваться, часто обусловлены поспешным обобщением данных (информации) в определенных ситуациях, и в таких случаях критике подвергается не сама технология, а ее использование без убедительных научных доказательств.

Итак, рамочное правовое регулирование технологий искусственного интеллекта затрагивает документы, не имеющие обязательной юридической силы (I). Вместе с тем исследованию подлежат также пределы правового регулирования, оправдывающие применение правовых инструментов, имеющих обязательную юридическую силу (II).

**I. Правовые инструменты контроля за использованием технологий ИИ и рисками, связанными с их применением (не имеющие обязательной юридической силы)**

### **1. Определение искусственного интеллекта**

Определение понятия ИИ — довольно сложная задача, которую тем не менее необходимо выполнить. Стоит отметить, что на сегодняшний день не существует универсального понятия ИИ, принятого глобальным научным сообществом.

По мнению Комиссии по развитию французского языка, мы имеем дело с «междисциплинарной теоретической и практической областью, целью которой является понимание механизмов познания и размышления, а также их имитация аппаратным и программным обеспечением с целью оказания помощи или подмены человеческой деятельности».

На официальном сайте Совета Европы ИИ описывается как «совокупность научных достижений, теорий и методов, целью которых является имитация машиной когнитивных способностей человека».

ОЭСР дает такое определение понятию ИИ: «Система искусственного интеллекта (или система ИИ) — это автоматизированная система, которая для достижения определенного набора целей, формулируемых человеком, способна осуществлять прогнозы, формулировать рекомендации или принимать решения, влияющие на состояние реальной или виртуальной среды».

Объединенный исследовательский центр (ОИЦ) Европейской комиссии подготовил 97-страничный метаанализ всех имеющихся определений ИИ с целью выработки универсального, базового определения<sup>7</sup>. Можно сделать краткий вывод, что все зависит от той точки зрения, которая берется за основу.

Для формулировки определения ИИ в широком смысле можно исходить из наук с ярко выраженной когнитивной составляющей, например кибернетики. Сужение фокуса рассмотрения приближает нас к определению ИИ, относящемуся к информатике и ее различным

<sup>4</sup> URL: <https://www.coe.int/en/web/artificial-intelligence/ai-and-control-of-covid-19-coronavirus>.

<sup>5</sup> URL: <https://www.theverge.com/2020/8/17/21372045/uk-a-level-results-algorithm-biased-coronavirus-covid-19-pandemic-university-applications>.

<sup>6</sup> URL: <https://www.en24news.com/2020/11/football-smart-camera-confuses-the-ball-and-the-linesmans-skull.html>.

<sup>7</sup> URL: <https://op.europa.eu/en/publication-detail/-/publication/6cc0f1b6-59dd-11ea-8b81-01aa75ed71a1/language-en>.

подходам к программированию рассуждений (символические или соединительные модели).

С 2010 г. определение совершенно точно и недвусмысленно относится к машинному обучению, которое включает в себя в качестве составных элементов глубокие нейронные сети и так называемое глубокое обучение<sup>8</sup>.

Именно поэтому обсуждение проблем искусственного интеллекта всецело зависит от фокуса рассмотрения: имеем ли мы в виду программы «зеркального отражения» мыслительной деятельности или автоматический поиск повторов в большой массе данных? Каждый раз мы имеем дело с разными явлениями как с точки зрения техники, так и с точки зрения возможности применения технологий.

Таким образом, действие алгоритмической механики возможно при определенных технологических условиях и весьма ограничено в других, поэтому не стоит опасаться возможной неопределенности в отношении развития технологий искусственного интеллекта.

## 2. Подлежащие оценке риски использования технологий ИИ

Широкое применение так называемого статистического подхода, основанного на использовании данных и прогностических возможностях машинного обучения, привело к универсализации использования алгоритмов машинного обучения в широком спектре человеческой деятельности — торговле (профилирование рекламы), логистике, промышленности, а также в тех секторах, где довольно сложно автоматизировать человеческую деятельность, например в сфере здравоохранения (медицинская диагностика) или правосудия.

В целом направления использования ИИ можно разделить на следующие категории, уточнив, что некоторые приложения ИИ могут быть комплексными (сочетать в себе несколько категорий):

- анализ данных в целях выявления нетипичной деятельности или повторов — здесь речь идет о возможности обнаружения, например, случаев мошенничества (банковская сфера, функционирование различных служб) или установления поведенческих профилей лиц для различных целей (например, для рекламного таргетинга);
- интерпретация данных, основанная на структурированной или неструктурированной информации (системы анализа

медицинских рентгеновских снимков, системы распознавания лиц);

- прогнозирование в целях оценки развития той или иной ситуации — цель здесь состоит в том, чтобы дать рекомендации относительно контента (например, в новостных лентах социальных сетей) или установить вероятность возникновения того или иного события (например, используя балльную систему) в различных приложениях по оценке риска;
- взаимодействие с людьми — чат-боты или инструменты, которые имеют возможность реагировать на голос или действия своих пользователей;
- физическое взаимодействие с окружающей средой путем обработки информации от различных датчиков передачи данных — речь идет о роботизированных системах (автономных транспортных средствах), взаимодействующих с человеком, или иных роботизированных системах;
- взаимодействие между алгоритмическими системами — предположение о том, что алгоритмы взаимодействуют автономно (смарт-контракты).

Новые направления использования технологий ИИ зачастую возникают из вышеперечисленных направлений, не предполагая при этом возникновения идентичных по содержанию рисков, связанных с их использованием.

Должны ли мы, исходя из этого, обеспечивать такую же правовую основу для использования технологий ИИ, машинного обучения, в целях оказания помощи инвалидам или расширения человеческих возможностей?

Использование незрячими людьми шлема с распознаванием лиц для того, чтобы идентифицировать имена людей, к которым они обращаются, возможно, не следует рассматривать наравне с технологиями, функционал которых заключается в установлении связи между мозгом человека и компьютером для улучшения когнитивных способностей.

Если рассматривать проблему в более широком ракурсе, можно отметить, что в обществе растет обеспокоенность по поводу возможности массовой слежки, дискриминационных практик, применение которых может быть вызвано активным внедрением автоматизированных систем принятия решений, а также ослаблением демократических начал и верховенства права.

<sup>8</sup> Cardon D., Cointet J-P., Mazières A. La revanche des neurones, Réseaux 2018/5 n°211. 2018. P. 173 à 220.

Безусловно, нам необходимы объективные и научно обоснованные рамки функционирования и оценки технологий ИИ. При этом основой должно являться законодательство.

### 3. Действующая нормативная основа функционирования технологий ИИ в рамках Совета Европы (документы, не имеющие обязательной юридической силы)

Цель деятельности Совета Европы состоит в оказании содействия 47 государствам — членам данной региональной международной организации в более эффективной защите прав человека в соответствии с ценностями демократии и верховенства права.

Уже в 1981 г. Совет Европы принял знаковые международно-правовые документы в области регулирования цифровых технологий — Конвенцию 108 о защите данных, которая впоследствии вдохновила государства — члены Европейского Союза на принятие Будапештской конвенции о борьбе с киберпреступностью 2001 г.

Этот опыт фактически легализует деятельность Совета Европы, который в координации с Европейским Союзом, ОЭСР, Организацией Объединенных Наций, а также ЮНЕСКО и ВОЗ должен принять меры по разработке правовых норм в отношении ИИ.

Подавляющее большинство структурных подразделений и органов Совета Европы в последние годы анализировали последствия использования ИИ в отдельных секторах экономики. На основе таких исследований были подготовлены доклады, а также приняты документы мягкого права, не имеющие обязательной юридической силы. Вот список таких документов, который, однако, не претендует на то, чтобы считаться исчерпывающим:

- Рекомендация Уполномоченного по правам человека «Расшифровка МА: 10 шагов к защите прав человека»;
- Европейская этическая хартия по использованию искусственного интеллекта в судебных системах и их окружении (СЕРЕ)<sup>9</sup>;
- Руководящие принципы по искусственному интеллекту и защите данных (Т-РД)<sup>10</sup>;
- рекомендация Рабочей группы по искусственному интеллекту разработать международно-правовой документ, который станет основой для создания национального

законодательства об автоматизированном обучении;

- серия из семи докладов Парламентской ассамблеи Совета Европы (ПАСЕ), посвященных демократии, рискам возникновения дискриминационных практик, занятости в системе правосудия, занятости в сфере здравоохранения, влиянию на рынок труда, правовым аспектам автономных транспортных средств и взаимодействию между человеческим мозгом и компьютерами<sup>9</sup>.

Здесь идет речь о вертикальном, специализированном подходе Совета Европы. В так называемом горизонтальном аспекте в сентябре 2019 г. Комитет министров поручил Специальному комитету по искусственному интеллекту (САНАИ) изучить целесообразность разработки правового документа на основе консультаций с участием заинтересованных сторон, а также структурные элементы такого документа.

Мандат САНАИ заключается в работе над нормативным регулированием, дополняемым не имеющими обязательной юридической силы инструментами, вместе способными сформировать общие принципы использования ИИ.

## II. Юридические обязательные инструменты регулирования искусственного интеллекта и рисков, возникающих в связи с его использованием

### 1. Этические аспекты использования технологий машинного обучения

Предметы ведения, предоставленные Комитетом министров Специальному комитету, являются ярким подтверждением недостаточности инструментов саморегулирования, в том числе с помощью этических норм. Проверка, проведенная Секретариатом Совета Европы, выявила (по состоянию на ноябрь 2020 г.) более 360 документов (не имеющих обязательной юридической силы), в которых затрагивается тема использования ИИ<sup>10</sup>.

Большая часть документов приходится на период между 2017 и 2018 гг. Однако с 2019 г. наблюдается заметный спад в отношении их принятия по ряду причин. Несмотря на то, что интерес к этическому подходу неуклонно растет, его трансформация в целое научное

<sup>9</sup> URL : <https://pace.coe.int/en/pages/artificial-intelligence>.

<sup>10</sup> URL : <https://www.coe.int/en/web/artificial-intelligence/national-initiatives>.

направление (биоэтику) не дает конкретных ответов на возникающие вопросы.

Метаанализ, проведенный ЕТН (Цюрих), указывает на то, что многие принципы использования ИИ многозначны<sup>11</sup> (прозрачность, воспринимаемая с точки зрения частного оператора, может не совпадать с прозрачностью, рассматриваемой с точки зрения неправительственной организации, защищающей неприкосновенность частной жизни).

Кроме того, установлению доверительных отношений между пользователями препятствует отсутствие механизма обеспечения соблюдения этических принципов. В случае несоблюдения правил не следует никаких негативных последствий, и мы видим, как в случае с регулированием Интернета, что «вес» крупных операторов влияет на то, чтобы их предложение было структурировано в соответствии с уровнем налагаемых на них регуляторных ограничений.

В этом контексте и перед лицом потенциальных нарушений прав человека в результате использования технологий ИИ, на государства — члены Совета Европы возложено соблюдение двух обязательств: соблюдать эти права и обеспечивать их соблюдение, в частности посредством действий частных операторов.

## 2. На пути к юридически обязывающей основе, соразмерной рискам использования ИИ?

Принцип обязательного нормативного регулирования ИИ представляется сегодня возможным решением проблем, связанных с применением этой технологии. Необходимость создания доверительных отношений лежит в основе обеспокоенности Европейского Союза, поэтому Европейский парламент опубликовал ряд документов.

В феврале 2020 г. Белая книга Европейской комиссии по ИИ, озаглавленная «Европейский подход, основанный на совершенстве и доверии», установила те же вехи, предусмотрев регулирование для так называемых высокорисковых приложений. Специальный комитет по искусственному интеллекту Совета Европы в своем докладе о ходе работы в сентябре 2020 г. также рассматривает идею о разработке конкретных нормативных мер использования ИИ и оценке рисков.

Идея регулирующих органов заключается в том, чтобы не налагать новое регулятивное бремя на предпринимателей. Такая дифференциация уже существует в других отраслях промышленности, таких как фармацевтика, и привела не к замедлению инноваций или экономического роста, а, напротив, к созданию доверительных отношений с потребителями.

Малые и средние предприятия, безусловно, будут в наибольшей степени затронуты новыми обязательными мерами, но их применение будет необходимо только в том случае, если предприниматели решат работать в секторе деятельности, требующем особых мер предосторожности.

Что касается машинного обучения, то риски, которые необходимо предотвратить, в настоящее время хорошо известны и находят свое отражение в документах: дискриминационная практика, риски для автономии процесса принятия решений, посягательства на демократические процессы, свобода выражения мнений и достоинство.

В отчетах ПАСЕ делается акцент на большинство из них, предлагается, как и Европейскому парламенту, разработать обязательный для исполнения правовой акт, охватывающий не только предотвращение этих рисков путем проведения предварительных исследований, но и снижение этих рисков путем принятия соответствующих мер, таких как маркировка или сертификация. Добавим к этому, увязывая предлагаемые меры с простым принципом предосторожности, можно, таким образом, ввести мораторий на некоторые виды применения машинных алгоритмов в период оценки их воздействия.

Данные аспекты могут стать частью содержания возможной конвенции (рамочной конвенции), основанной на правах человека, демократии и верховенстве права и являющейся частью более широкого глобального механизма регулирования цифровизации, как это рекомендуется в докладе Организации Объединенных Наций, озаглавленном «Эпоха цифровой взаимозависимости».

## 3. От принципов к практике

Однако введение в действие такого правового инструмента будет нелегким делом. Действительно, не следует предполагать, что таким правовым инструментом будет «вол-

<sup>11</sup> Jobin A., Ienca M., Vayena E. The global landscape of AI ethics guidelines, *Nature Machine Intelligence* 1. Septembre 2019. P. 389—399.

шебная пуля», способная предотвратить любое замедление в использовании ИИ в долгосрочной перспективе. Воплощение принципов высокого уровня в конкретный процесс разработки компьютерных приложений оправдывает промежуточные меры, которые могут исходить из нескольких источников.

Первым из этих источников станет продолжение разработки не имеющих обязательной юридической силы документов в форме рекомендаций, руководящих принципов и руководств, которые могут быть легко адаптированы к новым вопросам и потребностям заинтересованных сторон.

Они будут опираться на имеющий обязательную юридическую силу правовой акт об основах использования ИИ и смогут обеспечить практическое воплощение общих принципов. Например, предварительные оценки воздействия на права человека могут быть

результатом работы комитета или многопрофильной группы экспертов, ориентированной на разработчиков технологий (приложений) ИИ.

Второй из этих источников может носить сквозной характер и трансформировать в форме поддающихся проверке стандартов все обязательные и необязательные меры. Такие организации, как IEEE, ISO или CEN уже работают над такими инструментами с целью обеспечения всей отрасли оперативным переводом правовых документов в результате коллективных и многосторонних обсуждений.

Механизмы оценки и мониторинга, подобные уже проведенным Советом Европы в других областях, могли бы дополнить эту панораму, чтобы предоставить государствам помощь в осуществлении указанных мер в виде постоянного обзора воздействия всей системы регулирования.

## **БИБЛИОГРАФИЯ**

1. Anders G. "The Politics of Technology" by E. Jolly. — Michalon, *The Common Good*, 2017.
2. Basdevant A., Mignard J.-P. *L'empire des Données, Essai sur la société, les algorithmes et la loi* — Don Quichotte, 2018.
3. Bensamoun A. (dir.), Loiseau G. (dir.) *Artificial Intelligence Law // LGDJ*. 2019.
4. Bourcier D., Hasset P., Roquilly C. (dir.) *Droit et intelligence artificielle — Une révolution de la connaissance juridique*. — Romillat, coll. *Droit et technologies*, 2000.
5. Cardon D. *What algorithms dream about — Our life in the era of big data*. — Threshold, call. Republic of Ideas, 2015.
6. Cardone D., Coint Zh-P., Mazier A. *Revenge of neurons*. — Rezo. — 2018/5. — № 211. — 2018. — P. 173—220.
7. David M., Sauviat C. *Artificial Intelligence, La Nouvelle Barbarie* — Roche ed., 2019.
8. Garapon A., Lassegue J. *Digital Justice — Revolutionary Graphics and the Anthropology of the Divide*. — PUF, 2018.
9. Jensen P. *Why Society Doesn't Allow Itself to Be Equated* — Threshold, 2018.
10. Julia L. *L'intelligence artificielle n'existe pas*. — 1st ed. — 2019.
11. Menecœur Y. *L'intelligence artificielle en procès: coll.* — Macro Droit Micro Droit. — Bruylant, 2020.
12. Morozov E. *To solve everything, click here — Aberration of the technological solution*. — FYO ed., 2014.
13. O'Neill C. *Weapons of Mathematical Destruction: How Big Data Increases Inequality and Threatens Democracy*. — 1st ed. — 2016.
14. Pearl, J., Mackenzie D. *The Why Book — A New Science of Cause and Effect*. — Basic Books, 2018.
15. Sadin E. "Artificial Intelligence" or "Magical Life". — Ed. L'Écappé. — 2018.
16. Stigler (ed.). *Digital Truth, Research and Higher Education in the Digital Age*. — FIP ed. — 2018.
17. Supiot A. *La gouvernance par les nombres — Course at the Collège de France (2012—2014)*. — Fayard, coll. — Poids et mesures du monde, 2015.
18. Tarissan F. *Au coeur des réseaux — Des sciences au citoyen*. — Éditions Le Pommier, 2019.