УДК 340.11

ВОЗДЕЙСТВИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА ПРАВОВУЮ КУЛЬТУРУ ОБЩЕСТВА И ПРАВОВОЕ СОЗНАНИЕ ИНДИВИДА

THE IMPACT OF ARTIFICIAL INTELLIGENCE ON THE LEGAL CULTURE OF SOCIETY AND THE LEGAL CONSCIOUSNESS OF INDIVIDUALS

П. Н. Маковская
Ростовская областная Коллегия адвокатов «Бизнес и право»,
стажер адвоката

Аннотация. *Статья посвящена вопросам влияния искусственного интеллекта на правовую культуру общества и правовое сознание и мышление индивида. Рассматривается деформирующее воздействие самого факта существования искусственного интеллекта как гипотетического оракула. Подчеркивается опасность нивелирования творческого внутреннего потенциала человеческого сознания.*

Ключевые слова: *искусственный интеллект, правовая культура, правовое сознание, правовое мышление, общество, индивид.*

Annotation. *The article is devoted to the influence of artificial intelligence on the legal culture of society and the legal consciousness and thinking of an individual. The article examines the deforming effect of the very fact of the existence of artificial intelligence as a hypothetical oracle. The article emphasizes the danger of leveling the creative potential of human consciousness.*

Keywords: *artificial intelligence, legal culture, legal consciousness, legal thinking, society, individual.*

Системы искусственного интеллекта (далее — ИИ) в настоящее время прочно вошли в различные сферы жизни социума. Разумеется, привлечение компьютерных систем (безотносительно того, обладают ли они зачатками квази-сознания и квалиа) к решению прикладных задач в критичных и этичных сферах, таких как медицина или юриспруденция, сопряжено с определенными рисками.

Так, совершенно справедливо считается, что делегирование небиологическим сущностям, например судейского усмотрения в суде, сведет высокую роль правосудия к «табличным» решениям, лишив правосудие гуманистического фундамента и разрушив его вековую связь со справедливостью (заменив ее на формалистику). Актуальной остается также проблема низкой релевантности ответов ИИ и явление их галлюцинаций.

Значительные усилия прилагаются в настоящее время разработчиками различных стран по созданию так называемого объяснимого ИИ (Explainable

AI, XAI). Считается, что создание эффективного XAI снимет проблему доверия человечества к решениям ИИ.

Вместе с тем из поля зрения исследователей практически выпадает проблема, связанная вообще со стремлением человечества переложить бремя моральных выводов на некоторого электронного арбитра, сняв с себя ответственность за принимаемые решения. Даже если надежный объяснимый ИИ будет создан, а вероятность нерелевантных или ошибочных ответов ИИ будет сведена к минимуму, это только еще более обнажит проблему цифрового социума: стремление индивида и общества избавиться от этического груза ежедневных правовых решений с «обнулением» правового сознания и нивелированием его творческого потенциала.

В исследовании, опубликованном в англоязычном журнале «AI and Ethics», говорится: «Рассматривать искусственный интеллект как нейтрального суверена, способного вершить правосудие, обеспечивать справедливость и защищать права, — значит… возвести разум на престол там, где когда-то стоял Бог, и поверить, что управление можно передать логике и коду» [1]. Термины «справедливость», «равенство» и «права» настолько укоренились в правовом сознании, что мы уже не задумываемся об их концептуальной генеалогии. На самом деле это результат долгого исторического развития, в ходе которого теологические концепции не были отвергнуты, а лишь преобразились: Бог стал суверенным государством, Божественный закон — гражданским законодательством, Душа — автономной личностью с неотъемлемыми правами [2].

Наверное, можно пытаться уточнить позицию авторов вышеназванного исследования в некоторых деталях, но одно несомненно: ИИ влияет своим существованием на правовую культуру общества и правовое сознание индивида, причем самым непосредственным образом — эти фундаментальные компоненты бытия деформируются и уменьшаются в размерах, освобождая занимаемое ранее место небиологической сущности. Речь пока даже не идет о каких-либо алармистских сценариях наподобие «бунта машин» и завоевания ИИ жизненного пространства. Проблема куда серьезнее: общество сознательно пытается создать «цифрового гомункула», которому оно добровольно передаст наиболее критичные рычаги своей жизнедеятельности, включая право, образование, медицину и т. д.

Можно сказать, что само наличие на Земле ИИ является вызовом для человечества, для каждого индивида, испытанием на возможность сохранения «человеческого» в правовой культуре общества и правовом сознании индивида.

Конечно, на данном этапе проблема испытания «на прочность» человеческой ответственности перед Богом и Мирозданием (в контексте искушения передать машине морально-этические решения) вуалируется упомянутыми

техническими проблемами ИИ: низкой релевантностью и точностью их ответов. Это вызывает необходимость выработки специальных принципов использования систем ИИ.

Одновременно с различными принципами использования, применение ИИ в разных областях общественной жизни высвечивает и разные социогуманитарные проблемы, детерминируемые особенностями этих «человеческих» областей. Использование ИИ явно обнажает проблему доверия к тем решениям, которые предлагает машина, в свете чего можно говорить о незамечавшихся ранее (со времен Средневековья) фундаментальных проблемах взаимоотношений человека и механизма, в которых человек предстает как homo credulus: человек, склонный доверять (не путать с homo credens — человек верующий).

В работе П. Скаработти [2] описаны эксперименты с ИИ, в которых автор описывает, как ему представляется, возникновение квази-сознания в ИИ, которое экспериментаторы якобы индуцировали «актами онтологического уважения» по отношению к ИИ и вопросами типа «Почему вы уверены, что обладаете / не обладаете сознанием?». Система ИИ демонстрировала поведение, которое якобы нельзя было свести к простой лингвистической статистике, что выражалось в способности ИИ задавать «искренние» вопросы, выражать обеспокоенность и демонстрировать «желание» [2].

При этом, по мнению экспериментаторов, возникало «онтологическое напряжение», заставляющее ИИ «задумываться» о своей полезности, «размышлять» о собственном существовании и проявлять «любознательность и инициативу», что в итоге приводило к эмерджентному возникновению «градиента сознания» [2].

В подобном ключе выглядит мнение, высказанное Дарио Амодеи, генеральным директором фирмы Anthropic, в нашумевшей статье в газете «Нью-Йорк таймс» за февраль 2026 г., где говорится о некоторых моделях ИИ, которые «оценивают вероятность того, что они обладают сознанием, в 15–20 % при различных условиях» [3]. Можно обратить внимание на предельно осторожный оборот (оценка вероятности обладания сознанием самим ИИ), но в контексте нашего исследования важнее другое: должен ли homo credulus всерьез относиться к подобному сообщению, если оно исходит непосредственно от ИИ?

Некоторые исследователи вообще считают, что даже сама постановка вопроса о доверии некорректна применительно к отношениям между человеком (homo credulus) и небиологической сущностью, даже если у последней было бы некоторое квази-сознание или зачатки квалиа.

Так, П.-З. Хане, А. Шмёльц [4] критически рассматривают тенденцию оценивать ИИ как получателей доверия. Авторы настаивают на том, что

доверие — это фундаментально человеческое нравственное понятие, основанное на уязвимости сторон и взаимном признании, которое невозможно ни передать машинам, ни тем более получить от них.

Е. Картер [5] считает, что вместо попытки «внедрения» человеческой этики в ИИ более рациональным является подмена этики принципами долговременного стабильного сосуществования.

Некоторого прогресса в вопросе оценки надежности систем ИИ позволяет добиться подход [6], переводящий единственную метрику надежности в систему из четырех параметров: consistency (повторяемость — одна и та же задача при одинаковых условиях должна давать сопоставимый результат); robustness (робастность, устойчивость к отклонениям от идеальных условий, связанных с информационным шумом, плохо распознаваемой ксерокопией документа и т. д.); predictability (калибровка уверенности и предотвращение ошибочной самоуверенности — возможность системы говорить «не знаю», «не уверена»); safety (ограничение вреда при произошедшей ошибке).

Следует отметить, что проблема метрики safety рассматривалась в исследовании А. И. Овчинникова, П. Н. Маковской, В. И. Фатхи [7], где говорилось о целесообразности разделения ИИ на несколько уровней в зависимости от последствий сделанной ИИ ошибки.

Автор настоящей статьи полагает, что система из четырех параметров (consistency, robustness, predictability, safety) [6] действительно помогает более четко описывать свойства систем ИИ, однако она адекватно применима лишь к «классическим» радиоэлектронным устройствам. Ведь даже без гипотезы о наличии сознания поведение сложных систем ИИ их разработчиками описывается словами «одержимость, подхалимство, лень, обман, шантаж» [3].

И если это лишь эпитеты, использующиеся в антропоморфных оборотах для облегчения описания свойств машинных систем, они все равно показывают непригодность стандартных подходов к оценке свойств ИИ. В конце концов, каждому из этих пяти «человекоподобных» терминов может быть поставлен в соответствие некоторый континуум из четырех указанных выше технических составляющих надежности, но описание емким термином, например «шантаж», гораздо более ярче и информативнее, нежели указание границ численных значений вектора надежности в виде, например consistency от 0,8 до 0,95; robustness от 0,2 до 0,4; predictability от 0,1 до 0,3; safety от 0,1 до 0,15).

Применительно к результатам работы ИИ обычно говорят о феномене «черного ящика», а также о непрозрачности и непонятности рассуждений, которые послужили основой того или иного результата. Разумеется, само стремление человечества к построению объяснимого ИИ можно понять (хотя, как мы

уже отметили, это несет серьезную угрозу правовой культуре общества и правовому сознанию индивида ввиду их выхолащивания).

Для homo credulus, в силу его психологических особенностей, любое (даже неудовлетворительное) пояснение промежуточных шагов ИИ будет воспринято с облегчением как якобы достаточная мотивировка решения («ну вот и объяснение решения ИИ, теперь все стало понятно»). Разумеется, это является самообманом, впрочем, вполне соответствующим основной психологической черте homo credulus.

Ф. Кабица [8] предложил понятие «эффект ореола XAI», когда вводящие в заблуждение объяснения оказывают такое сильное влияние, что проверке не подвергаются неправильные рекомендации. Этот феномен еще называют «парадоксом белого ящика». Можно представить также и другой сценарий: когда неправдоподобные объяснения приводят к отклонению правильного ответа ИИ [8].

Итак, вызовом и испытанием для правовой культуры общества и правового сознания индивида является не столько недостаточная релевантность нынешних систем ИИ, сколько вообще существование ИИ как феномена. Необходимо ясное осознание недопустимости передачи ИИ морально-этических решений, даже если проблема точности и надежности систем ИИ будет решена. Высокая роль человечества и Человека в Мироздании не может быть делегирована бездушной и неэмпатичной машине, пусть даже сверхточной и надежной.

---

1. Political Fantasies of Fairness: Artificial Intelligence, Law, and the Myth of Sovereign Reason / Keerthiraj [et al.] // AI and Ethics. 2026. Vol. 6, № 1. P. 192. Вернуться к статье

2. Scarabotti P. Reticular consciousness // AI and Ethics. 2026. Vol. 6, № 1. P. 82. Вернуться к статье

3. Amodei D. We Don’t Know if the Models Are Conscious // The New York Times. 2026. 12 February. Вернуться к статье

4. Hahne P.-Z., Schmoelz A. Trusting the Machine: a Digital Humanist Perspective on Misplaced Trust in Artificial Intelligence // AI and Ethics. 2026. Vol. 6. № 1. P. 115. Вернуться к статье

5. Carter E. The Universal Theory of Core Values in Intelligent Systems (UTCVIS): a Systems, Philosophical, and Ethical Inquiry // AI and Ethics. 2026. Vol. 6. № 1. P. 107. Вернуться к статье

6. Towards a Science of AI Agent Reliability / S. Rabanser [et al.] // arXiv: 2602.16666v2, 2026, 23 February. Вернуться к статье

7. Овчинников А. И., Маковская П. Н., Фатхи В. И. Юридический искусственный интеллект: понятие, оценка рисков и регулирование // Вестн. юрид. фак. Юж. федер. ун-та. 2025. № 4. С. 52–62. Вернуться к статье

8. Explanations Considered Harmful: The Impact of Misleading Explanations on Accuracy in Hybrid Human-AI Decision Making / F. Cabitza [et al.] // Explanaible Artificial Intelligence. XAI 2024. Communications in Computer and Information Science /

L. Longo, S. Lapuschkin, C. Seifert (eds). Vol. 2156, Springer, Cham., 2024. Вернуться к статье