

Научная статья

УДК 174 + 177

DOI 10.25205/2541-7517-2024-22-4-5-24

Законы робототехники Азимова: мировоззренческие основы и когнитивные искажения

Илья Алексеевич Суров

Национальный исследовательский университет ИТМО

Санкт-Петербург, Россия

ilya.a.surov@itmo.ru, <https://orcid.org/0000-0001-5690-7507>

Аннотация

«Законы робототехники» Айзека Азимова пользуются исключительным авторитетом и являются своего рода заповедями для многих сторон этически-философского дискурса вокруг искусственного интеллекта. Практическое применение этих «законов», однако, сталкивается с принципиальными трудностями, причины которых исследованы в этой статье. В частности, первый «закон» о недопущении вреда человеку не уточняет, об ущербе кому идёт речь; второй «закон» о подчинении робота человеку не уточняет, кому из претендентов на управление роботом следует отдать предпочтение. На элементарных примерах показано, что при наличии конфликта интересов оба этих абстрактно-гуманистических «закона» теряют смысл и являются жизненно несостоятельными. Кроме того, даже в отсутствие такого конфликта «недопущение вреда» требует от робота предвидения последствий тех или иных действий в сколь угодно далёком будущем, однако возможность такого прогноза противоречит законам физики и биологии. Наконец, показано, что формулировки всех трёх «законов», очевидно абсурдные для неодушевлённых инструментов, предполагают наличие у роботов свободы действия, ошибочно наделяя их субъектностью. Этот ложный антропоморфизм, характерный для современных обсуждений искусственного интеллекта, закрепился в устойчивых языковых формах. Система выявленных искажений представляет собой чуждую традиционным культурам когнитивно-мировоззренческую матрицу, угрожающую безопасности общества на переходе к новым техноукладам.

Ключевые слова

Законы робототехники, искусственный интеллект, Азимов, абстрактный гуманизм, антропоморфизм, детерминизм, субъектность, когнитивные искажения, этика, право

Благодарности

Работа выполнена при поддержке фонда Потанина.

Для цитирования

Суров И. А. Законы робототехники Азимова: мировоззренческие основы и когнитивные искажения // Сибирский философский журнал. 2024. Т. 22, № 4. С. 5–24. DOI 10.25205/2541-7517-2024-22-4-5-24

© Суров И. А., 2024

Asimov's laws of robotics: conceptual foundations and cognitive fallacies

Илья А. Суров

ITMO University, St. Petersburg, Russia

ilya.a.surov@itmo.ru, <https://orcid.org/0000-0001-5690-7507>

Abstract

Asimov's laws of robotics provide rare common ground, a kind of commandments for all sides of ethical discourse on artificial intelligence. Practical implementation of these «laws», however, faces fundamental problems which are studied in this work. In particular, the first «law» on the prevention of harm to humans does not specify, the harm to whom is meant; the second «law» on the subordination of a robot to a human does not specify, which of the subjects willing to control this robot must be preferred. Elementary examples show that with ubiquitous conflicts of interest, both these abstractly-humanistic laws are untenable. Even in the absence of such conflicts, the «prevention of harm» requires a robot to foresee the consequences of certain actions in the arbitrarily distant future, which contradicts the laws of physics and biology. Finally, it is shown that formulations of all three laws, obviously ridiculous in application to inanimate instruments, assume that robots have freedom of action, fallaciously endowing them with subjectness. Such misplaced anthropomorphism, typical for the ongoing discussions of artificial intelligence, has a foothold in stable language forms. A bundle of the revealed fallacies makes up an erroneous conceptual system, threatening social security in transition to future technological lifestyles.

Keywords

Robotics, artificial intelligence, Asimov, abstract humanism, anthropomorphism, determinism, subjectness, cognitive distortion, ethics, law

Acknowledgements

The work was supported by the Potanin Foundation.

For citation

Surov I. A. Asimov's Laws of Robotics: conceptual foundations and cognitive fallacies // *Siberian Philosophical Journal*. 2024. Vol. 22, no. 4. C. 5–24. DOI 10.25205/2541-7517-2024-22-4-5-24

1. Введение

Искусственный интеллект и кибернетические технологии – важная часть создаваемого техноуклада, которая во многом определит будущее человечества. Нормы использования этих технологий формируются на основе различных научно-профессиональных, культурно-религиозных и нравственно-мировоззренческих позиций (см.: [Карпов и др., 2018; Введенская, 2019; Разин, 2019; Попова, 2020; Алексеева, 2024; Дронова, 2024]). В силу новизны рассматриваемой проблематики имеющееся разнообразие точек зрения, подходов и сторон практически лишено общепризнанных авторитетов. Редким исключением является Айзек Азимов – американский писатель-фантаст, работавший над этикой человекоподобных машин начиная с 40-х годов прошлого века¹. В сборнике рассказов под названием «Я, робот», дополнявшемся более поздними произведениями (см.: [Азимов, 1994]), Азимов рассматривает аспекты взаимодействия человека и робота, многие

¹ Полное имя – Исаак Юдович Азимов (1920, РСФСР – 1992, США). См.: «Азимов, Айзек» (ред. 5 января 2025). *Википедия. Свободная энциклопедия*. URL: https://ru.wikipedia.org/wiki/Азимов,_Айзек (дата обращения: 01.02.2025).

из которых становятся актуальны лишь сейчас, а некоторые остаются фантастической и поныне.

Творческое наследие Азимова особенно ценно тем, что основные его выводы в отношении робототехнического, искусственного интеллекта обобщены в трёх кратких тезисах, называемых «законами робототехники» (далее – «законы»). Азимов отмечал, что эти «законы» – едва ли не самое важное, что он написал за всю свою жизнь, тогда как большая часть его (более пятисот) остальных произведений во многом посвящена исследованию действия этих «законов» [Asimov, 1950]².

Первый закон: робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинён вред (A robot may not injure a human being or, through inaction, allow a human being to come to harm).

Второй закон: робот должен повиноваться всем приказам, которые даёт человек, кроме тех случаев, когда эти приказы противоречат первому закону (A robot must obey the orders given by human beings except where such orders would conflict with the First Law).

Третий закон: робот должен заботиться о своей безопасности в той мере, в которой это не противоречит первому или второму законам (A robot must protect its own existence as long as such protection does not conflict with the First or Second Law).

Иными словами, робот должен быть безопасен для человека, управляем, а также по возможности сохраняться в работоспособном состоянии. Азимов связывает эти правила с нормами общечеловеческой этики: первый «закон» подобен библейской «любви к ближнему»; второй «закон» соответствует «порядочности» человека, следующего обычаям культуры, авторитетам и распоряжениям правительства; третий «закон» соответствует инстинкту самосохранения (см.: [Азимов, 1994, с. 519]).

В современной терминологии рассматриваемые Азимовым робототехнические и кибернетические системы обозначаются термином «искусственный интеллект». Положение о том, что такой искусственный интеллект (ИИ) должен находиться под контролем человека и не должен причинять ему вред, представляется бесспорным. Того же можно пожелать для любого нашего орудия, инструмента и приспособления, в связи с чем согласно Азимову его «законы» распространяются (хотя и не всегда успешно) на все такие инструменты (см.: [Латиева, 2019]). В связи с этим «законы», мысленно опробованные в широком спектре жизненных обстоятельств, пользуются у исследователей гуманитарных аспектов кибернетики и техники в целом исключительным авторитетом. Подобно библейским «не убий» и «не укради» «законы» фактически стали своего рода заповедями, общим местом этического дискурса в области ИИ, робототехники и кибернетики (см.: [Clarke, 1993; Алимова, Каменская, 2019]).

В настоящей статье показано, что это первое впечатление обманчиво и на практике ведёт к весьма нежелательным последствиям. В разделе 2 рассмотрен ряд простых примеров, проявляющих неприменимость «законов» в кон-

² В действительности «законы Азимова» сформулированы еще в 1940 году его наставником Джозефом Кэмпбеллом – главным редактором американского научно-популярного журнала «Astounding Science Fiction», в котором «законы» были впервые опубликованы в марте 1942 года [Алимова, Каменская, 2019].

кретных жизненных обстоятельствах. В разделе 3 показано, что приведённая формулировка «законов» противоречит основам традиционных культур правового регулирования ИИ. В разделах 4 и 5 рассмотрены причины незаметности этих противоречий, а также их мировоззренческие основы. В заключительном разделе 6 отмечены элементы языковой практики, внимание к которым необходимо для предупреждения возможных проблем развития информационных технологий.

2. Абстрактный гуманизм

Попробуем применить «законы» на деле. Рассмотрим беспилотный робот-автомобиль, стоящий на городской улице в режиме ожидания. ИИ такой машины обязан распознавать людей, животных, дорожную разметку и транспортные средства, прогнозировать их будущее движение, вероятности столкновений и других аварийных ситуаций. Наблюдая за дорожной ситуацией в фоновом режиме, система распознаёт на противоположной стороне улицы двух человек, идущих друг на друга, глядя в свои телефоны. Вероятность их столкновения оценивается как высокая, причём в результате возможно травмоопасное падение людей на лёд.

Согласно первому «закону», робот обязан не допустить травм и другого ущерба в результате потенциального столкновения. Для этого ИИ включает двигатель машины и, не нарушая правил дорожного движения, направляет её к точке вероятного столкновения, громко сигнала. Столкновения и травм удаётся избежать, но в результате этого манёвра на дороге возникает затор. Кто-то опаздывает на встречу, звуки гудков мешают чьей-то беседе. Пробка отображается на онлайн-картах навигаторов других машин города, которые перестраивают свои маршруты в объезд. Больше время в пути повышает транспортные расходы и выбросы выхлопных газов, которые портят экологию и здоровье горожан.

Удалось ли роботу подчиниться первому «закону» о ненанесении ущерба человеку? Очевидно, что нет. Как видно из намеченных выше последствий (цепочку которых можно продолжать), в сложившейся ситуации исключить ущерб людям невозможно, так что первый «закон» робототехники оказывается невыполним.

В связи с этим его формулировку можно было бы модифицировать, обязав робота не устранять, а минимизировать общий ущерб³. Для этого, однако, потребовалось бы заранее определить метрики такого ущерба для всевозможных последствий, что также невозможно. Кроме того, при этом потребовалось бы, например, сравнить ущерб от травм столкнувшихся лиц и ущерб здоровью граждан (а также их будущих поколений) от загрязнения окружающей среды, и многое другое; ясно, что соответствующие численные оценки также не представляются возможными. В рассказе «Улики» рассматривается лишь вполне однозначный случай, когда для спасения людей роботом может быть убит сумасшедший преступник (см.: [Азимов, 1994, с. 520]).

³ В поздних рассказах, согласно которым роботы участвуют в управлении государствами и цивилизациями, Азимов дополняет три приведённых закона «нулевым», который устанавливает для роботов приоритет «интересов человечества» над ущербом для отдельных людей (см.: [Азимов, 1994, с. 559]).

С минимизацией ущерба связан ещё один немаловажный вопрос: об ущербе кому идёт речь? Продажа вредной еды, алкоголя и наркотиков наносит ущерб потребителю, одновременно принося пользу производителям и продавцам. На такие продажи робот вполне мог бы влиять в рамках обычного законодательства – например, загородив своим корпусом рекламу магазина или вход в него. Как ущерб всех вовлечённых сторон будет учтён в метриках, которыми при этом будет руководствоваться робот согласно первому «закону»?

Эти ситуации – не какой-то исключительный, специально выдуманный сюжет наподобие дилеммы вагонетки, в которой «законы» Азимова также заходят в тупик (см.: [Persson, Hedlund, 2024]); проблема «стороны» возникает во всех коммерческих, политических, личных отношениях обособленных субъектов, в которых конфликт интересов является нормой. Человек хотел бы приобрести долговечную и надёжную вещь, но тогда производитель не сможет продать ему следующую модель; пациент хотел бы выздороветь, но если это произойдёт, то «биг-фарма» потеряет клиента; успех одной компании на рынке означает банкротство другой; сделанный на заводе беспилотник или бомба спасёт одних людей и уничтожит других (см.: [Шиббаева, Холова, 2018]). Чьё благополучие будет охранять робот по первому «закону»? Чьим приказам он будет подчиняться по второму? Эти фундаментальные неопределённости, замечаемые на удивление редко (см.: [McCauley, 2007; Bueno and Jankowski, 2024; Алексеева, 2024]), указывают на полную жизненную несостоятельность «законов».

Азимов вполне осознавал возможность военного применения роботов одними людьми против других (рассказ «Давайте объединимся»), однако прямое противоречие с «законами» при этом как бы не замечено. Эта проблема затронута лишь в заключительном рассказе цикла «Яко помниши его...», где выяснено, что на практике следование «законам» требовало бы от роботов собственной оценки целесообразности полученных приказов, количественного сравнения ущербов, причиняемых их действиями разным людям, а также предпочтения одних приказов другим⁴. При этом из нескольких противоречивых приказов, согласно Азимову, выполнять следует тот, который отдан более достойным человеком – по уму, характеру и знаниям – независимо от его физических параметров (см.: [Азимов, 1994, с. 619])⁵.

Большинство же (хронологически более ранних) сюжетов сборника [Азимов, 1994] обходят вопрос конфликта интересов стороной. Выведенные в них и общеизвестные ныне формулировки «законов» подразумевают, что противоречий отмеченного типа в мире попросту нет; в этом идеальном мире как человек, так и «совершенный робот» может быть за всё хорошее против всего плохого. Как видно из приведённых примеров, этот абстрактный гуманизм есть грубое

⁴ При этом отмечено, что такие «дискриминирующие» суждения роботов нарушали бы права человека, как они понимаются в западной культуре (см.: [Азимов, 1994, с. 598]).

⁵ В сюжете показано, что эта логика ведёт роботов к выводу о том, что лучшими из людей являются именно они. Соответственно этому все три «закона» переориентируются на роботов, превосходящих человека по интеллекту и порядочности (в смысле следования «законам»). Эти «совершенные роботы» и становятся приоритетным объектом защиты и наиболее авторитетным источником приказов для других роботов [Там же].

и опасное искажение действительности, химера гуманизма реального (см.: [Водолазов, 2013]).

3. Субъектность: оглашения и умолчания

Правовое регулирование искусственного интеллекта – острая проблема современности, требующая согласования технической и гуманитарной мысли. Возможные решения этой проблемы в пределе сходятся к одной из следующих альтернатив (см.: [Попова, 2021]).

1. ИИ есть объект права наравне с другими видами имущества, инструментов и информационной собственности.

2. ИИ есть субъект права, обладающий некоторой частью прав животных и человека, адаптированных под специфику так называемого «электронного лица»⁶.

Второй вариант рассмотрен Азимовым в рассказе «Двухсотлетний человек», по сюжету которого один из роботов, изначально запрограммированный как камердинер, дворецкий и горничная, со временем обрёл способность к непредсказуемому, свободному и творческому поведению; специалисты объяснили это тем, что сложность «позитронных связей» в искусственном мозгу делала возможным лишь приближённое решение соответствующих уравнений. Робот начал заниматься художественной резьбой по дереву, а получаемый в результате доход был оформлен на его собственное имя «Эндрю» (однокоренное латинскому «андроид»). Поначалу «Эндрю» тратил деньги на нужды своего владельца, минимизируя его ущерб согласно первому «закону»; по прошествии некоторого времени «Эндрю» «захотел» приобрести себе свободу аналогично самовыкупу рабов у своих хозяев. Последовала «великая битва за права роботов» (против «болезни человечества» – «страха человека перед роботом», «инстинктивной антипатии» консерваторов, «недоступных доводов рассудка» – цитаты из текста), включая право ношения одежды и обладания органическим телом, право на свободу, жизнь и защиту от человека; частью этой борьбы стало воздействие на общественное мнение посредством научно-популярных лекций и серии судебных прецедентов. Одно из промежуточных решений сформулировано так: «Нельзя отказывать в свободе тому, кто обладает сознанием, развитым в степени достаточной, чтобы воспринимать понятие свободы и желать её» [Азимов, 1994, с. 632]. Финальным актом является решение всемирного суда, впервые в истории признавшее робота человеком (см.: [Азимов, 1994, с. 667]).

За исключением этого позднего рассказа, изобразившего вариант далёкого будущего, большинство сюжетов Азимова явно поддерживает первую из двух представленных позиций: поскольку движения робота определяются работой заложенных в него программ и не подвержены какому-либо произволу, ответственности за свои действия робот нести в принципе не может. В силу отсутствия правосубъектности роботы, как и любой другой инструмент, являются собственностью человека или производящей их компании, которая и отвечает за их дей-

⁶ См. также: [Владимирова, 2021] или Зорькин В.Д. «Право и вызовы искусственного интеллекта» (27.06.2024). *Российская газета*. URL: <https://rg.ru/2024/06/27/pravo-i-vyzovy-iskusstvennogo-intellekta.html> (дата обращения: 01.02.2025).

ствия (см.: [Азимов, 1994, с. 510]). Сами же роботы собственностью обладать не могут, не пользуются свободой, неприкосновенностью и другими правами человека⁷.

Эта позиция Азимова, однако, противоречит неявному содержанию «законов», которое нетрудно увидеть в приведённых формулировках. Чтобы выявить это, посмотрим, как «законы» звучали бы на русском языке в приложении к другим инструментам человека.

Первый «закон»: кран / лопата / пистолет не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинён вред.

Второй «закон»: кран / лопата / пистолет должен повиноваться приказам человека кроме тех случаев, когда эти приказы противоречат Первому закону.

Третий «закон»: кран / лопата / пистолет должен заботиться о своей безопасности в той мере, в какой это не противоречит первому и второму «законам».

Нелепость этих предложений, столь же грубая в англоязычном оригинале, режет слух любому носителю языка. Дело в том, что как в английском, так и в русском миропонимании неодушевлённый предмет – будь то трактор, лопата или камень на дороге – не может «причинять вред», «действовать» или «бездействовать», «повиноваться» или «не повиноваться», что-то «допускать» и о ком-то «заботиться». Эти и другие подобные выражения применимы исключительно к одушевлённым существам, способным к произвольному, свободному поведению. Соответственно, если в списке приведённых предложений заменить кран / лопату / пистолет на человека или животного, то отмеченная нелепость исчезает. Являясь субъектами свободного поведения, человек и собака (в меру своего понимания) несут ответственность за совершённые действия и могут быть заслуженно наказаны. Наказание же инструмента за что бы то ни было, напротив, совершенно неадекватно природе происходящего (см.: [Melnikova, Surov, 2023]) – хотя настоящий виновник, вероятно, был бы рад такому «правосудию».

Таким образом, формулировки «законов» подразумевают, что их объектом является одушевлённая сущность, способная к произвольному поведению. Это наделяет робота правосубъектностью в соответствии уже не с первым, а со вторым вариантом из отмеченных в начале этого раздела – в противоречии с буквальным содержанием большинства рассказов Азимова. В отмеченном выше рассказе «Двухсотлетний человек», написанном спустя 25 лет после основной серии «Я, робот», эта логика одушевления робота представлена в следующем диалоге:

– *А ведь робот не неодушевлённый предмет. И он не животное. ...*

– *Он не имеет понятия о свободе. Он робот.*

– *Папа, ты его не знаешь. Он прочёл все книги в нашей библиотеке. Я не знаю, что он ощущает внутри, но ведь я не знаю, что и ты ощущаешь в душе. Попробуй поговорить с ним, и ты убедишься, что самые различные абстрактные понятия он воспринимает как ты и я, а что ещё имеет значение? ...*

⁷ Право человека на личную неприкосновенность получает важное следствие при необходимости определить, является данная сущность роботом или человеком. Как показано Азимовым, такое определение требует обследования внутреннего устройства этой сущности, например с помощью рентгена. Принуждение к такому обследованию, однако, нарушает права человека, за которого может выдавать себя робот. В результате определить, является ли эта сущность роботом или человеком, в рамках правового поля становится невозможно (см.: [Азимов, 1994, с. 523–525]).

Имея в виду внимание Азимова к тонкостям формулировок, эта противоречивость оглашений и умолчаний вряд ли является случайной⁸. Сюжеты сборника [Азимов, 1994] указывают на то, что такую двойственную позицию автор избрал намеренно. Согласно этим сюжетам, для создателей роботов и прочей интеллектуальной элиты ясно, что роботы – снабжённые необходимой механикой искусственные вычислительные машины, запрограммированные ими для работы в определённых целях. Остальной же, менее образованной части общества осознать разницу между живым существом и искусной имитацией трудно и, возможно (с точки зрения Азимова), даже и ни к чему; соответствующая сцена представлена в вышеприведённой цитате и многих других сюжетах цикла. В рассказе «Лучший друг мальчика», например, ребёнка не убеждают объяснения отца о разнице между роботом и настоящей собакой (см.: [Азимов, 1994, с. 14]). Для этой части населения вполне годится «детский вариант» объяснения механики роботов как одушевлённых существ, «заботящихся», «чувствующих», «действующих», «не допускающих» и «повинующихся» приказам человека; именно это и достигается тройкой «законов»⁹. Несмотря на свою фундаментальную неадекватность, правдоподобность такого восприятия имеет глубокие психологические основания.

4. Антропоморфизм и механоморфизм

Для безопасности любого организма необходима предсказуемость внешней среды и последствий своих собственных действий. Такая предсказуемость требует от каждого индивида понимания касающихся его процессов, то есть знания управляющих ими законов. Смена времён года и полёт стрелы, например, следуют простым закономерностям и потому хорошо предсказуемы, что позволяет безопасно использовать эти процессы в своих целях.

Достичь такого понимания, однако, удаётся не всегда; движение машин на дороге, динамика рыночных цен, поступки людей, животных и многое другое не описываются никакой простой формулой. Для таких процессов желательно отыскать управляющего ими субъекта и выяснить его намерения, цели, умонастроения и чувства. Как и знание законов физики в предыдущих примерах, это «психологическое» знание делает такие процессы более предсказуемыми и безопасными, хотя полной предсказуемости в этих случаях достичь так и не удаётся.

В доиндустриальную эпоху понимание по первому, хорошо предсказуемому варианту было выражено в традиционных канонах строительства, земледелия, медицины, мореплавания и военного дела, за рамками которых лежал непознанный мир непредсказуемых и потенциально опасных процессов природы. Соглас-

⁸ По ходу сюжета рассказа «Улики» происходит диалог робота с судебным инспектором. В ходе этого разговора робот замечает, что утверждение о владении некоторой собственностью юридически однозначно указывает на то, что речь идёт не о роботе, а о человеке; соответственно, эта сущность обладает соответствующими правами, защищаемыми законом (см.: [Азимов, 1994, с. 525]).

⁹ В нескольких рассказах руководители единственного на планете (американского) конструкторского бюро «U.S. Robots and Mechanical Men, Inc.» отмечают, что государственных чиновников и широкую публику желательно не посвящать в детали их работы, чтобы не создавать лишних трудностей (см.: [Азимов, 1994, с. 567]). Стержневая и авторитетная героиня цикла – робопсихолог Сьюзен Кэлвин – в разговоре с этими руководителями замечает: «Вы делите человечество на категории и большую часть из них презираете или не принимаете в расчёт» [Азимов, 1994, с. 588].

но второму сценарию, управление этими процессами относилось на счёт одушевлённых сил со своими чувствами, настроением, характером и волей; фактически любой сложный процесс за пределами немногих ремесленных канонов считался осмысленным поведением некоторого живого субъекта, в принципе подобного человеку¹⁰. В течение тысячелетий жизни человека в естественной среде этот мысленный приём уподобления других одушевлённых существей человеку – антропоморфизм – работал вполне эффективно (см.: [Guthrie, 1997]).

Появление роботов и так называемого искусственного интеллекта принципиально изменило окружающую среду человека. Создателям ИИ ясно, что это не более чем системы электронных «шестерёнок», подобные огромным часовым механизмам. Независимо от их сложности действие таких механизмов полностью предопределено их памятью и входными данными; двигаясь по программе в обратном направлении, причину каждого выданного ими результата можно установить однозначно (при современной сложности таких систем, однако, технических возможностей для этого не всегда хватает даже у их разработчиков)¹¹. Сторонний же пользователь видит лишь сложное, практически непредсказуемое человекоподобное действие – будь то движение танцующего робота от Boston Dynamics или поток текста в диалоговом окне Chat GPT. Отлаженный тысячелетиями когнитивный алгоритм распознаёт такую сущность как живую, мысленно наделяя её собственными намерениями, целями, чувствами и волей (см.: [Placani, 2024]). Этот механизм аналогичен психологической проекции, посредством которой дети наделяют кукол чертами характера по своему образу и подобию (см.: [Борзых, 2019]). Такая игра полезна в детстве как подготовительный этап к взаимодействию с реальными людьми; имитационный ИИ, напротив, чаще используется в обратном направлении замещения человека искусственным суррогатом. При этом чем совершенней имитация, тем привлекательнее становится это направление трансгуманизма».

Подобно личностям людей, животных и других организмов, мнимая личность робота или диалогового ИИ наивно наделяется всеми её обычными возможностями: переживать, заботиться, подчиняться и не подчиняться, действовать и бездействовать по своему разумению и свободной воле. Именно это восприятие развёрнуто в творчестве Азимова, согласно которому роботы глубоко понимают происходящее, проникательно и стратегически мыслят, тонко переживают, общаются на языке высокого стиля, заботятся о людях, принимают ответственные решения и показывают образцы благородства; в изображении Азимова эти качества делают даже «несовершенных» роботов более привлекательным, чем большинство человеческих персонажей. Многократно обкатывается сюжет, в котором робот неотличим от человека, а человек в итоге оказывается роботом, помогая читателю привыкнуть к одинаковому восприятию тех и других.

¹⁰ Устанавливая наличие психики во всё большем спектре индивидуальных и коллективных организмов (см.: [Lyop и др., 2021; Levin, 2022]), современная биоинформатика показывает, что при таком одушевлении допускалось намного меньше ошибок, чем это было принято считать в материалистических философиях XVIII–XX веков.

¹¹ В рассказе «Яко помниши его...» создатель наиболее совершенного человекоподобного робота не сразу напоминает себе, что заслуга этого робота в решении сложной проблемы «не больше, чем у логарифмической линейки» [Азимов, 1994, с. 618].

В этом уравнении присутствует и обратная сторона антропоморфизма – механоморфизм (см.: [Sapota, 1986]), то есть восприятие человека как механической и электрохимической Машины с большой буквы (как она пишется в части рассказов). При этом намерения, чувства и волю человека следует считать такой же иллюзией, как у робота; принципиальное отличие людей состоит лишь в невоспроизводимой сложности органического тела и соответствующих биополей (см.: [Rubik, 2018]). В остальном робот и человек оказываются, по сути, одним и тем же, а отмеченное в предыдущем разделе противоречие разрешается само собой. Будучи – как и многие выдающиеся писатели – хорошим психологом, Азимов не пишет об этом прямо, позволяя читателю сознательно или подсознательно самому прийти к этому выводу.

5. Мировоззренческие основания

Привлекательность идеи разумного, «совершенного» человекоподобного робота имеет под собой не только житейско-психологические, но и глубинные мировоззренческие основы. На эти основы указывает то обстоятельство, что исполнение «законов» требует от робота предвидения последствий совершаемых им действий. При этом такое предвидение необходимо на неограниченно большой срок поскольку, как отмечает один из героев Азимова, «достижение кратковременных целей зачастую оборачивается долговременными неприятностями» [Азимов, 1994, с. 607], чему в повседневной жизни имеется сколько угодно примеров, когда отложенное следствие давно завершившихся действий меняет их оценку с положительной на отрицательную и обратно. Не допуская подобных ошибок, «законы» Азимова подразумевают, что в распоряжении роботов должна быть программа, позволяющая им прогнозировать сколь угодно отдалённые последствия доступных роботу движений.

Для наивно-научного мировоззрения, образцовым носителем которого был Азимов, возможность такого прогноза представляется вполне правдоподобной. Имея в распоряжении неограниченные вычислительные ресурсы, возможно, на ещё не открытых физических принципах – например, «платино-иридиевый, позитронный мозг» – принципиальных препятствий для решения этой задачи обычно не просматривается; и тем не менее они есть.

Получив высшее химическое образование в США на рубеже 1930–1940-х годов, Азимов был знаком с новыми на тот момент открытиями квантовой физики, вынужденно отказавшейся от классических представлений о полной предсказуемости природы. Как отмечается азимовскими робото-инженерами, квантовая неопределённость поведения частиц на субатомном уровне есть единственное принципиальное отличие человеческого мозга от «позитронного», действие которого полностью предопределено заранее; предполагается, что именно эта фундаментальная непредсказуемость отвечает за интуицию, творчество, спонтанность и свободу человека, отличающие его от машин (см.: [Азимов, 1994, с. 566]). Эта гипотеза подтверждается по ходу рассказа «Женская интуиция», согласно которому квантовая модификация «позитронного мозга» позволила новой версии робота интуитивно-иррациональным образом решить задачу о нахождении ближайших к Земле

обитаемых планет. В отличие от предыдущих моделей, эту машину нельзя было выключать; у неё появилось истинно эмоциональное «женское» поведение, а во всем, что «она» делала, чувствовался «скрытый смысл» (см.: [Азимов, 1994, с. 579]).

В этих описаниях Азимов весьма точно предвосхитил открытия современной нейробиологии, согласно которым в поведении живых организмов есть фундаментально непредсказуемая часть (см.: [Еськов, 2017; Еськов и др., 2022]), в принципе не поддающаяся алгоритмической имитации; в одной из теорий причиной этой неалгоритмичности является усиление квантовой неопределённости с атомно-молекулярного масштаба на макроскопический уровень поведения живых организмов (см.: [Jedlicka, 2017; Суров, 2023]). Как и у «интуитивного» робота Азимова, это усиление делает возможным заранее не предопределённое, свободное поведение человека, недоступное для сколь угодно сложного ИИ на обычных вычислительных принципах (см.: [Surov, Melnikova, 2024]). Усиленная на макроуровень квантовая неопределённость предоставляет место для свободы выбора и воли, произвольно-осмысленного поведения, переживаний от первого лица, сознания и чувств, ответственности и юридической правосубъектности, которые представляют собой единый неделимый блок [Там же]. Художественные картины об эмоциональном, сознательном, этическом и вместе с тем детерминированном, бесправном и безответственном роботе, разводя неотъемлемые стороны единого явления, уравнивают человека и робота с меньшим стрессом для читателя, по частям.

Помимо свободного человека или даже «квантового робота» в окружающей его среде действует множество других субъектов, поступки которых также принципиально непредсказуемы. Классическим примером является нематода *Caenorhabditis elegans*, нервная система которой из 302 нейронов полностью картирована в 1986 году, однако спрогнозировать на этой основе движение какого-либо отдельного живого образца до сих пор не удалось (см.: [Toyoshima et al., 2024]). То же самое относится к отдельным живым клеткам, в основе поведения которых лежат не только классические вычисления, но и неалгоритмизируемые квантовые процессы (см.: [Панов, 2014]). Аналогичной способностью усиления микроскопических флуктуаций обладают хаотические процессы, например в атмосфере, что накладывает принципиальные ограничения на горизонт их предсказуемости¹².

Динамика хаотических систем и спонтанное поведение живых организмов определяют будущее внешней для робота среды, а также её отклик на действия робота; в силу природы квантовой неопределённости ни то ни другое не может быть ни имитировано, ни предсказано. Соответственно, ни одно мышление ни на классических, ни на квантовых принципах не сможет предвидеть будущее так, как это предполагается «законами» Азимова¹³.

¹² По этой причине более-менее достоверный прогноз погоды ограничен двумя неделями, и увеличить этот срок не поможет никакая вычислительная мощность ни классических, ни квантовых компьютеров.

¹³ Математически это фундаментальное ограничение выражено в теореме Гёделя – Тьюринга, согласно которой для любой конечной программы можно предъявить утверждение, которое она не может подтвердить, ни опровергнуть. Эта теорема означает, что мир принципиально невычислим и не может быть имитирован никаким сколь угодно сложным алгоритмом (см.: [Панов, 2014]).

Таким образом, «законы робототехники» не просто трудны для осуществления на практике в силу невычислимости метрик ущерба, конфликтности субъектов управления или «неполноты информации», как можно было бы заключить на основе примеров из раздела 2; эти вымышленные «законы» (как и их многочисленные модификации [Clarke, 1994; Введенская, 2019]) находятся в глубинном противоречии с истинными законами природы, исключающими предопределённость и предсказуемость будущего; общая схема этих мировоззренческих искажений показана на рисунке ниже. Соответственно, идея «совершенного робота» (как и совершенного человека) с объективно оптимальным поведением в интересах «всеобщего блага» является физически несостоятельной.

Объективное, безусловное добро может иметь место как идеал, стремление к которому в целом свойственно природе человека; сам этот идеал, однако, существует только в силу отличия добра от зла и возможности свободного выбора между ними. Для робота ни такое различие, ни такой выбор недоступны, а значит идея всеобщего добра в «мире машин» осуществиться в принципе не может. Поэтому абсолютно упорядоченное, «объективно оптимальное» течение жизни на Земле под управлением «совершенного робота» – даже если бы оно допускалось законами природы – означало бы не триумф, а самоликвидацию человечества (см.: [Высочина, 2023]).



Схема мировоззренческих искажений, ведущих к идее «совершенного робота», потенциально превосходящего человека. Стрелками показаны логические следствия
 Scheme of conceptual fallacies behind the idea of a «perfect robot», potentially superior to a human being.
 Arrows show logical implications as explained in the text

6. Языковая матрица

На основе приведённых аргументов блогер N мог бы сообщить, что «все поступки, действия и чувства робота предопределены заложенной в него программой»; после этого могло бы идти пояснение, что «есть ли у робота чувства и что это такое вообще – вопрос до конца не изученный, но современные модели разговорного ИИ всё лучше понимают человека и реагируют на них всё более

естественно, демонстрируя высокий уровень эмпатии, соответствующий людям такого-то возраста». Получив подобное сообщение из ленты новостей, немногие читатели найдут время вникнуть в тонкости проблемы. Большинство запомнит главное: у робота есть поступки, действия и чувства, а также эмпатия и понимание – что, как показано выше, не соответствует действительности. Где произошла ошибка?

Читатель понял всё верно: словосочетание «высокое дерево», например, подразумевает, что дерево есть пространственный объект, занимающий много места по вертикальному направлению. Такого типа выводы – важная способность естественного мышления, лежащая в основе, например, чрезвычайно эффективного (по сравнению с ИИ) обучения детей языку безо всяких учебников просто путём слушания разговоров окружающих (см.: [Frensch and Rünger, 2003; Cleeremans et al., 2019]). Таким же образом указанный выше ложный вывод следует из приведённого в кавычках текста, в языковых оборотах которого («поступки, действия и чувства робота», «ИИ понимает» и «демонстрирует эмпатию») и содержится искажение. Тот факт, что в соседних словах сказано о предопределённости и научных затруднениях, дела не меняет; эта информация адресована сложному логическому мышлению, требующего умственных усилий и отсутствующего у читателя времени.

Подобным образом можно весьма квалифицированно рассуждать о принципах действия ИИ, объяснять его внутренне устройство и алгоритмику, обосновывать допустимость или недопустимость наделения ИИ правосубъектностью; но если при этом используются фразы «ИИ научился» (то есть научил себя сам), «ИИ принял верное / неверное решение», «ИИ ошибся / сделал правильный выбор», «ИИ поступил так или иначе», «ИИ настаивает / предпочитает / относится», «доверие к ИИ» и тому подобные, то такая письменная или устная речь – намеренно или не намеренно – навязывает адресату вывод о субъектности ИИ. Кто совершил поступок, принял верное решение или ошибся, тот и несёт ответственность – верный логический ход, автоматически достраиваемый подсознанием читателя. То же самое следует из льстящих самолюбию разработчиков обращений «мать / отец роботов», поскольку родиться может только живое, одушевлённое существо.

По такому же принципу действуют цитированные и многие другие рассказы Азимова, в которых роботы стремятся, действуют, подчиняются, чувствуют, любят и обижаются, страдают, хотя и наслаждаются. Как отмечено выше, Азимов использует эти выражения вполне намеренно, осознавая важность «составления фраз с учётом всех оттенков смысла» [Азимов, 1994, с. 643]. Встречая подобные языковые обороты во всевозможных обстоятельствах, неявное мышление – как при рассуждениях курящего о вреде курения – легко отсеивает логическую маскировку и воспринимает истинный авторский посыл в обход сознания читателя.

Таким образом, язык действует как когнитивная матрица, определяющая для своих носителей ходы мысли, возможные суждения и выводы в отношении ИИ, роботов и других кибернетических систем. Языковая матрица, сформированная трудами Азимова (и других подобных ему деятелей западной науки и культуры), противоречит фундаментальным законам природы (раздел 5) и ведёт своих

последователей к катастрофам; в связи с этим культурам, желающим сохраниться в будущем, от этой матрицы целесообразно было бы отстроиться.

Для такой острой необходимости прежде всего внимательное отношение к письменной и устной речи (в которой поставить кавычки жестом не всегда уместно), не допускающее использования человекоподобных оборотов по отношению к роботу (и наоборот, раздел 4). Часть таких выражений имеет более корректные технические аналоги, которые вполне могут войти в обиходный язык. Например, вместо «машинных рассуждений» можно говорить о машинном выводе, вычислениях, анализе или переработке данных; другие выражения, вроде «жизни / смерти роботов», «этике роботов», «машинной интуиции», «машинного принятия решений», «сознания / чувств / эмоций ИИ» и тому подобные, напротив, следовало бы исключить как опасно искажающие реальность, несмотря на их метафорическую ценность и популярную эффектность (см.: [Батурин, Полубинская, 2022; Placani, 2024]). Для их замены возможно создание новых понятий без антропоморфной окраски, способствующих безопасному развитию информационного техноклада.

7. Заключение

Отмеченные свойства языка обсуждаются редко, а если и обсуждаются, то как вторичные по отношению к созданию технологических реалий «на земле». История социокультурного и научно-технического развития, однако, раз за разом показывает несостоятельность такого подхода; «бытие», вне всякого сомнения, определяет сознание, но в действительности это лишь половина правды, поскольку сознание определяет бытие ничуть не хуже (а может быть и лучше – если «в начале» действительно было слово). Супруги или партнёры, пол или гендер, отец или родитель номер два – разные названия определяют разные практики, допустимые и недопустимые в отношении именуемой сущности. Эти неявные определения задают мировоззренческие основания и этические ориентиры для решения проблем современности, формируя будущий облик человечества (см.: [Гребенникова, 2018; Латиева, 2019; Нигоматуллина, 2022]).

В этом плане рассказы сборника [Азимов, 1994] направлены на переосмысление понятия «робот» и наполнение его новым содержанием. Роботы Азимова под водительством «законов» заменяют образ «чудовища Франкенштейна» на образ друзей и помощников человека в контрасте с кошмарными сюжетами М. Шелли и К. Чапека (см.: [Маслова, 2018; Кружалова, 2023]). Современные прецеденты наделения роботов собственными интересами, гражданством, элементами правосубъектности, авторским и другими правами человека обусловлены именно этим смысловым сдвигом¹⁴.

В предыдущих разделах показано, что для этого сдвига – как и для вызываемого произведениями Шелли, Чапека, Замятина, Хаксли ужаса перед роботами (см.: [Маслова, 2018]) – оснований столько же, как для страха перед пистолетом, самолётом или беспилотным трактором. Всё это более или менее сложные механизмы,

¹⁴ См.: [Владимирова, 2021] или Зорькин В.Д. «Право и вызовы искусственного интеллекта» (27.06.2024). *Российская газета*. URL: <https://rg.ru/2024/06/27/pravo-i-vyzovy-iskusstvennogo-intellekta.html> (дата обращения: 01.02.2025).

которые по недостатку образования может одушевлять туземец-абориген; особенность современного ИИ состоит в том, что из-за его сложности в ловушке антропоморфизма оказалась большая часть человечества. В творчестве названных и других фантастов имеет место ложная эмоциональная раскачка публики, причём первый ход качелей – как и положено по законам жанра – выполнен в сторону страха.

Эффективность этой раскачки в течение более чем двух столетий зрима не только по отмеченным правовым прецедентам и законодательным инициативам. В этом же русле лежат многочисленные модификации «законов» Азимова и формируемые под их влиянием «концепции» (см.: [Батурин, Полубинская, 2022]) и «кодексы этики ИИ» как принципы его якобы самостоятельного поведения (см.: [Карпов и др., 2018; Алексеева, 2024]). Как и другие рассмотренные примеры, «этика ИИ» – столь же ложное понятие, как «этика ножей» или «этика консервных банок», однако именно это фундаментальное мировоззренческое искажение навязывает читателю Азимов под умело приготовленным научным соусом. Троянский конь этого искажения в виде «законов робототехники» распознан, однако зашёл уже достаточно далеко.

Список литературы

- Азимов А.** Миры Айзека Азимова. Книга первая. Совершенный робот. М.: Полярис, 1994. 668 с.
- Алексеева И. Ю.** Этика искусственного интеллекта как прикладная этика // Философия и общество. 2024. № 3. С. 69–85.
- Алексеев А. П., Алексеева И. Ю.** Статус этических кодексов в этике искусственного интеллекта // Информационное общество. 2024. № 4. С. 43–49.
- Алимова С. З., Каменская И. Б.** Критическая рецепция цикла рассказов Айзека Азимова «Я, робот» // Формирование профессиональной компетентности филолога в поликультурной образовательной среде: Материалы II Международной научно-практической конференции / Под общ. ред. И. Б. Каменской. Евпатория: Индивидуальный предприниматель Корниенко Андрей Анатольевич, 2019. С. 23–26.
- Батурин, Ю. М., Полубинская С. В.** Искусственный интеллект: правовой статус или правовой режим? // Государство и право. 2022. № 10. С. 141–154.
- Борzych С. В.** Антропоморфизм ИИ // Журнал философских исследований. 2019. Т. 5. № 2. С. 2–9.
- Введенская Е. В.** Актуальные проблемы робоэтики // Научно-исследовательские публикации. 2019. № 2019. С. 88–101.
- Владимирова З. В.** О возможных способах наделяния искусственного интеллекта правосубъектностью // Вестник магистратуры. 2021. № 8. С. 50–55.
- Водолазов Г. Г.** Гуманизм абстрактный и гуманизм реальный // Гуманизм и современность: Материалы международной научно-образовательной конференции. Казань: Казанский (Приволжский) федеральный университет, 2013. С. 30–40.
- Высочина Д. С.** Повседневность прогнозируемого мира (Сборник рассказов Азимова «Я, робот») // Вестник Калужского университета. Серия 2. Исследования по филологии. 2023. № 3. С. 70–76.

- Гребенщикова Е. Г.** Этика как управление // Сибирский философский журнал. 2018. Т. 16. № 4. С. 135–146.
- Дронова Е. А.** Искусственный интеллект и социология морали // Свободная мысль. 2024. № 4. С. 55–62.
- Еськов В. М.** Конец определенности: хаос гомеостатических систем. Тула: ТППО, 2017. 595 с.
- Еськов В. М. и др.** Хаотический мозг // Сложность. Разум. Постнеклассика. 2022. № 2. С. 5–16.
- Карпов В. Э., Готовцев П. М., Ройзензон Г. В.** К вопросу об этике и системах искусственного интеллекта // Философия и общество. 2018. № 2. С. 84–105.
- Кружалова Л. В.** Проблемы взаимоотношений естественного и искусственного интеллектов (роботов) в произведениях научной фантастики (К. Чапек, И. Варшавский, А. Азимов и др.) // International journal of professional science. 2023. № 9. С. 5–13.
- Латиева А. Р.** Взаимодействие научно-фантастических концепций и социума // Проблемы современных социокультурных исследований / Под ред. Е.В. Хлыщевой, Р.Т. Алиева, В.Л. Тихоновой, О.И. Закутнова. Астрахань: Изд. дом «Астраханский университет», 2019. С. 77–81.
- Маслова К. К.** Трансформация образа «робот» от «механического подобия» к образу «робот-человек» в произведениях (К. Чапека, Е. И. Замятина, О. Хаксли и А. Азимова) // Вестник славянских культур. 2018. Т. 49. С. 161–170.
- Нигматуллина Р. М.** Этика как онтология управления // Сибирский философский журнал. 2022. Т. 20. № 1. С. 15–25.
- Панов А. Д.** Технологическая сингулярность, теорема Пенроуза об искусственном интеллекте и квантовая природа сознания // Информационные технологии (Приложение к журналу). 2014. № 5. С. 1–32.
- Попова А. В.** Этические принципы взаимодействия с искусственным интеллектом как основа правового регулирования // Правовое государство: теория и практика. 2020. Т. 16. № 3. С. 34–43.
- Попова Н. Н.** Развитие технологий искусственного интеллекта в перспективах трансформации права: правосознание в проекции будущего // Вестник Московского государственного областного университета. Серия: Юриспруденция. 2021. № 3. С. 65–76.
- Разин А. В.** Этика искусственного интеллекта // Философия и общество. 2019. Т. 1. № 90. С. 57–93.
- Суров И. А.** Квантовая концепция свободы выбора // Ученые записки Института психологии РАН. 2023. Т. 3. № 4. С. 68–82.
- Шиббаева К. В., Холова Л. Н.** Три закона робототехники Айзека Азимова: к вопросу о гуманности применения смертоносных автономных систем вооружения на войне // Теология, философия, право. 2018. Т. 8. № 4. С. 28–41.
- Asimov I. I.** Robot. New York: Gnome Press, 1950.
- Bueno C. C., Jankowski S.** Judgement after Automation: Posthumanist Reflections on Asimov's Laws of Robotics // Journal of science fiction and philosophy. 2024. Vol. 7. P. 1–19.
- Caporael L. R.** Anthropomorphism and mechanomorphism: Two faces of the human machine // Computers in human behavior. 1986. Vol. 2. No. 3. P. 215–234.

- Clarke R.** Asimov's laws of robotics: implications for information technology. Part 1 // Computer. 1993. Vol. 26. No. 12. P. 53–61.
- Clarke R.** Asimov's laws of robotics: Implications for information technology. Part 2 // Computer. 1994. Vol. 27. No. 1. P. 57–66.
- Cleeremans A., Allakhverdov V. M., Kuvaldina M.** (eds.) Implicit learning: 50 years on. Abingdon, Oxon: Routledge, 2019. 262 p.
- Frensch P. A., Rüniger D.** Implicit Learning // Current Directions in Psychological Science. 2003. Vol. 12. No. 1. P. 13–18.
- Guthrie S. E.** Anthropomorphism: A definition and a theory // Anthropomorphism, anecdotes, and animals / Ed. R. W. Mitchell, N. S. Thomson, H. L. Miles. New York: State University of New York Press, 1997. P. 50–58.
- Jedlicka P.** Revisiting the quantum brain hypothesis: Toward quantum (neuro)biology? // Frontiers in Molecular Neuroscience. 2017. Vol. 10. P. 1–8.
- Levin M.** Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds // Frontiers in Systems Neuroscience. 2022. Vol. 16. P. 768201.
- Lyon P. и др.** Reframing cognition: getting down to biological basics // Philosophical Transactions of the Royal Society B: Biological Sciences. 2021. Vol. 376. No. 1820. P. 20190750.
- McCauley L.** AI Armageddon and the Three Laws of Robotics // Ethics and Information Technology. 2007. Vol. 9. No. 2. P. 153–164.
- Melnikova E., Surov I.** Legal Status of Artificial Intelligence from Quantum-Theoretic Perspective // BRICS Law Journal. 2023. Vol. 10. No. 4. P. 5–34.
- Persson E., Hedlund M.** The Trolley problem and Isaac Asimov's First law of robotics // Journal of science fiction and philosophy. 2024. Vol. 7. P. 1–21.
- Placani A.** Anthropomorphism in AI: hype and fallacy // AI Ethics. 2024. Vol. 4. No. 3. P. 691–698.
- Rubik B.** Artificial intelligence and the human biofield: New opportunities and challenges // Cosmos and history: The Journal of natural and social philosophy. 2018. Vol. 14. No. 1. P. 153–162.
- Surov I. A., Melnikova E. N.** Subjectness of intelligence: quantum–theoretic analysis and ethical perspective // Foundations of science. 2024. <https://doi.org/10.1007/s10699-024-09947-y>
- Toyoshima Y. и др.** Ensemble dynamics and information flow deduction from whole-brain imaging data // PLOS computational biology. 2024. Vol. 20. No. 3. E1011848.

References

- Asimov I. I.** Robot. New York: Gnome Press, 1950.
- Asimov I.** The Worlds of Isaac Asimov. The first book. The perfect robot. Moscow: Polaris, 1994. 668 p. (in Russian)
- Alekseyeva I. Yu.** Ethics of Artificial Intelligence as Applied Ethics // Philosophy and Society. 2024. No. 3. P. 69–85. (in Russian)
- Alekseev A. P., Alekseeva I. Yu.** The status of ethical codes in the ethics of artificial intelligence // Information Society. 2024. No. 4. P. 43–49. (in Russian)

- Alimova S. Z., Kamenskaya I. B.** Critical reception of the cycle of Isaac Asimov's short stories "I, robot" // Formation of professional competence of a philologist in a multicultural educational environment. Materials of the II International Scientific and Practical Conference. ed. I.B. Kamenskaya. Yevpatoria, 2019. P. 23–26. (in Russian)
- Baturin Y. M., Polubinskaya S. V.** Artificial intelligence: legal status or legal regime? // Gosudarstvo i pravo. 2022. No. 10. P. 141–154. (in Russian)
- Borzykh S.** Anthropomorphism of artificial intellect // Journal of philosophical research. Vol 5. No. 2. P. 2–9. (in Russian)
- Bueno P. P., Jankowski S.** Judgement after Automation: Posthumanist Reflections on Asimov's Laws of Robotics // Journal of Science Fiction and Philosophy. 2024. Vol. 7. P. 1–19.
- Caporael L. R.** Anthropomorphism and mechanomorphism: Two faces of the human machine // Computers in Human Behavior. 1986. Vol. 2. No. 3. P. 215–234.
- Clarke R.** Asimov's laws of robotics: implications for information technology. Part 1 // Computer. 1993. Vol. 26. No. 12. P. 53–61.
- Clarke R.** Asimov's laws of robotics: Implications for information technology. Part 2 // Computer. 1994. Vol. 27. No. 1. P. 57–66.
- Cleeremans A., Allakhverdiv V. M., Kuvaldina M.** (eds.) Implicit learning: 50 years on. Abingdon, Oxon: Routledge, 2019. 262 p.
- Dronova E. A.** Artificial intelligence and the sociology of morality // Free thought. 2024. No. 4. P. 55–62. (in Russian)
- Eskov V. M.** The End of certainty: the chaos of homeostatic systems. Tula: TPPO, 2017. 595 p. (in Russian)
- Eskov V. et al.** Chaotic brain // Complexity. Mind. Postnonclassical. 2022. No. 2. P. 5–16. (in Russian)
- Frensch P. A., Rüniger D.** Implicit Learning // Current Directions in Psychological Science. 2003. Vol. 12. No. 1. P. 13–18.
- Grebenshchikova E. G.** Ethics as governance // Siberian Philosophical Journal. 2018. Vol. 16. No. 4. P. 135–146. (in Russian)
- Guthrie S. E.** Anthropomorphism: A definition and a theory // Anthropomorphism, anecdotes, and animals / Ed. R. W. Mitchell, N. S. Thomson, H. L. Miles. State University of New York Press, 1997. P. 50–58.
- Karpov V. E., Gotovtsev P. M., Roizenzon G. V.** On the issue of ethics and artificial intelligence systems // Philosophy and Society. 2018. No. 2. P. 84–105. (in Russian)
- Kruzhalova L. V.** Problems of the relationship between natural and artificial intelligence (robots) in the works of science fiction (K. Chapek, I. Varshavsky, A. Asimov, etP.) // International Journal of Professional Science. 2023. No. 9. P. 5–13. (in Russian)
- Latieva A. R.** The interaction of science fiction concepts and society. In the collection Problems of Modern Socio-cultural Research. Ed. E.V. Hlyshheva, R.T. Aliev, V.L. Tihonova, O.I. Zakutnov. Astrahan': Astrakhan University Publishing House, 2019. P. 77–81. (in Russian)
- Levin M.** Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds // Frontiers in Systems Neuroscience. 2022. Vol. 16. P. 768201.

- Lyon P. et al.** Reframing cognition: getting down to biological basics // *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2021. Vol. 376. No. 1820. P. 20190750.
- Maslova K. K.** Transformation of the «robot's» image from «mechanical similarity» to the «robot-human» (K. Chapeka, E. I. Zamyatin, O. Huxley, A. Asimov) // *Bulletin of Slavic Cultures*. 2018. Vol. 49. P. 161–170. (in Russian)
- McCauley L.** AI Armageddon and the Three Laws of Robotics // *Ethics and Information Technology*. 2007. Vol. 9. No. 2. P. 153–164.
- Melnikova E., Surov I.** Legal Status of Artificial Intelligence from Quantum-Theoretic Perspective // *BRICS Law Journal*. 2023. Vol. 10. No. 4. P. 5–34.
- Nigomatullina R. M.** Ethics as an ontology of management // *Siberian Philosophical Journal*. 2022. Vol. 20. No. 1. P. 15–25. (in Russian)
- Panov A. D.** Technological singularity, Penrose's theorem on artificial intelligence and the quantum nature of consciousness // *Information Technologies (Supplementary)*. 2014. No. 5. P. 1–32. (in Russian)
- Popova A. V.** Ethical principles of interaction with artificial intelligence as a basis of legal regulation // *The rule of law: theory and practice*. 2020. Vol. 16. No. 3. P. 34–43. (in Russian)
- Popova N. N.** Development of artificial intelligence technologies in the prospects of transformation of law: legal consciousness in the projection of the future // *Bulletin of the Moscow State Regional University. Series: Jurisprudence*. 2021. No. 3. P. 65–76. (in Russian)
- Persson E., Hedlund M.** The Trolley Problem and Isaac Asimov's First Law of Robotics // *Journal of Science Fiction and Philosophy*. 2024. Vol. 7. P. 1–21.
- Placani A.** Anthropomorphism in AI: hype and fallacy // *AI Ethics*. 2024. Vol. 4. No. 3. P. 691–698.
- Razin A. V.** Ethics of artificial intelligence // *Philosophy and society*. 2019. Vol. 1. No. 90. P. 57–93. (in Russian)
- Rubik B.** Artificial intelligence and the human biofield: New Opportunities and Challenges // *Cosmos and History: The Journal of Natural and Social Philosophy*. 2018. Vol. 14. No. 1. P. 153–162.
- Shibaeva K. V., Khokhlova L. N.** Isaac Asimov's Three Laws of robotics: on the question of the humanity of the use of deadly autonomous weapons systems in war // *Theology Philosophy Law*. 2018. Vol. 8. No. 4. (in Russian)
- Surov I. A., Melnikova E. N.** Subjectness of Intelligence: Quantum-Theoretic Analysis and Ethical Perspective // *Foundations of Science*. 2024. <https://doi.org/10.1007/s10699-024-09947-y>
- Surov I. A.** Quantum Concept of Free Choice // *Proceedings of the Institute of Psychology of the Russian Academy of Sciences*. 2023, Vol. 3. No. 4. P. 68–82. (in Russian)
- Toyoshima Y. et al.** Ensemble dynamics and information flow deduction from whole-brain imaging data // *PLOS Computational Biology*. 2024. Vol. 20. No. 3. P. E1011848.
- Vladimirova Z. V.** On possible ways of endowing artificial intelligence with legal personality // *Bulletin of the Magistracy*. 2021. No. 8. P. 50–55. (in Russian)

- Vodolazov G. G.** Humanism abstract and humanism real // Proceedings of the international scientific and educational conference «Humanism and Modernity» Kazan: Kazanskij (Privolzhskij) federal'nyj universitet, 2013. P. 30–40. (in Russian)
- Vvedenskaya E. V.** Actual problems of roboetics // Scientific research. 2019. № 2019. P. 88–101. (in Russian)
- Vysochina D. S.** The everyday life of the predicted world (collection of short stories by A. Asimov «I, robot») // Bulletin of the Kaluga University. Series 2. Studies in philology. 2023. No. 3. P. 70–76. (in Russian)

Информация об авторе

Илья Алексеевич Суров, кандидат физ.-мат. наук
старший научный сотрудник, Университет ИТМО

Information about the Author

Ilya A. Surov, Candidate of Sciences (Physics and Mathematics),
Senior Researcher, ITMO University

*Статья поступила в редколлегию 18.12.2024;
одобрена после рецензирования 24.02.2025; принята к публикации 27.02.2025*

*The article was submitted 18.12.2024;
approved after reviewing 24.02.2025; accepted for publication 27.02.2025*