

Gati A.L.,  
Suleymanov J.I.\*

DOI: 10.25108/2304-1730-1749.iolr.2023.72.49-66

UDC: 343.1

### Will a homunculus be able to conduct a criminal trial?

**Abstract:** The problems and trends in the use of artificial intelligence in criminal proceedings are considered.

The concepts, risks, system and structure of AI, existing points of view are explored.

**Key words:** artificial intelligence; criminal process; intelligence; risks.

### References

1. Abulkhanova-Slavskaya K.A. Strategy of life. Moscow, Mysl' Publ., 1991, 299 p. (in Russian).
2. An American drone decided to kill its operator during testing. Available at: URL: <https://ru.oxu.az/world/737463> (in Russian).
3. Ananyev B.G., Stepanova E.I. Development of psychophysiological functions of adults. Moscow, Pedagogika Publ., 1972, 246 p. (in Russian).
4. Biryukov P.N. Artificial intelligence and "predictive justice": foreign experience // LexRus-sica. - 2019. - No. 11(156). - P. 79-87 (in Russian).
5. Large psychological dictionary. Comp. by Mescherekov B.G. Moscow, Olma-Pres Publ., 2004. Available at: <https://vocabulary.ru/termin/intellekt.htm> (in Russian).
6. Wertheimer M. Psychology of productive thinking. - Moscow, Progress Publ., 1987, 336 (in Russian).
7. Hasanova Sh.Sh., Suleymanov J.I. Artificial intelligence and legal proceedings: problems and trends // Juridical Sciences and Education. - 2022. - No. 66. - P. 35-50. (in Russian).
8. Guilford J. Three sides of intelligence in the collection: Psychology of thinking. - Moscow, Progress Publ., 1965. - P. 433-456 (in Russian).
9. Golovin S.Yu. Dictionary of a practical psychologist. Minsk, Harvest Publ., 1998, 800 p.
10. Druzhinin V.N. Psychology of General Abilities. S. Petersburg, Piter Publ., 2007, 368 p. (in Russian).
11. Artificial intelligence spoke at a debate in Oxford, presenting arguments for and against itself. EverCare. 23 Dec. 2021. Available at: <https://evercare.ru/news/iskusstvenny-intellekt-vystupil-na-debatakh-v-oksforde-privodya-dovody-za-i-protiv-seby> (in Russian).
12. Artificial intelligence was used in court in China. Available at: [www.russiaedu.ru](http://www.russiaedu.ru) (in Russian).

---

\*Gati Alekper Lachin oglu – a dissertator of the National Academy of Aviation (Azerbaijan).  
E-mail: alakpar.l.gati@gmail.com

Suleymanov Javanshir Islam oglu – Doctor of Juridical Sciences, Professor (Azerbaijan).  
E-mail: mopi\_sid@yahoo.com

13. Köhler V. Some tasks of Gestalt psychology // Reader on the history of psychology. – Moscow, Moscow University Publ., 1980. - P. 102-120 (in Russian).
14. Kim L.M., Boretsky O.M. Al-Farabi's concept of intelligence and modernity//Vestnik of KazNU. Series Philosophy, anthropology, political sciences. 2016. No. 4(58). – P. 73-78 (in Russian).
15. China could become the first in the world to introduce artificial intelligence into the legal system. Available at: <https://www.inkstonenews.com/tech/china-may-become-worlds-first-bring-ai-legal-system/article/3125009> (in Russian).
16. Kostrikina I.S. The relationship between style and productive characteristics of intellectual activity in individuals with high IQ values. PhD in Psychology Diss. Moscow, Institute of Psychology RAS Publ., 2001, 207 p. (in Russian).
17. Corsini R., Auerbach A. Psychological Encyclopedia. S. Petersburg, Piter Publ., 2006. Available at: <https://vocabulary.ru/termin/intellekt-iskusstvennyi.html> (in Russian).
18. Kramarenko V.Yu. Intelligence and levels of its development. PhD in Philosophy Diss. Moscow, 1983, 230 p. (in Russian).
19. Oxford Explanatory Dictionary of Psychology / Ed. by A. Reber. S. Petersburg, 2002. Available at: <https://www.psyoffice.ru/6-487-iskustvenyi-intelekt.htm> (in Russian).
20. Penrose Roger. The new mind of the king. About computers, thinking and the laws of physics. Moscow, Editorial URSS, 2003. Available at: <http://vekordija.narod.ru/R-PENRO1.PDF> (in Russian).
21. Revonsuo Antti. Psychology of consciousness. –S. Petersburg, Piter Publ., 2013. Available at: <https://vocabulary.ru/termin/intellekt-iskusstvennyi.html> (in Russian).
22. A Microsoft robot turned into a racist and foul-mouthed person overnight. BBC News. March 24, 2016. Available at: [www.bbc.com/russian/society/2016/03/160324\\_tay\\_ai\\_racism](http://www.bbc.com/russian/society/2016/03/160324_tay_ai_racism) (in Russian).
23. Rubinshtein S.L. Problems of general psychology. Moscow, Pedagogika Publ., 1973., 424 p. (in Russian).
24. Handbook of general psychology. Intelligence. Available at: [https://spravochnick.ru/psihologiya/psihicheskie\\_svoystva\\_lichnosti/intellekt/](https://spravochnick.ru/psihologiya/psihicheskie_svoystva_lichnosti/intellekt/) (in Russian).
25. Suleymanov J.I., Shiraliyeva S.J. Problems of ethics and psychology when using AI in criminal proceedings // Juridical Sciences and Education. 2022. No. 69. – P. 140-147. (in Russian).
26. Tikhomirov O.K. Psychology of thinking. Moscow, MSU Publ., 1984, 272 p. (in Russian).
27. Criminal Procedure Code of the Republic of Azerbaijan as of 17.02.2023. Available at: [https://continent-online.com/Document/?doc\\_id=30420280](https://continent-online.com/Document/?doc_id=30420280) (in Russian).
28. Mind. Brain. Intelligence. Similarities and differences. Available at: <https://homogenes.ru/um-razum-intellekt-sxodstva-i-razlichiya-etix-ponyatij/> (in Russian).
29. Shapar V.B. The latest psychological dictionary. 44th edition. Rostov on Don, Fenix Publ. Available at: <http://ponjatija.ru/taxonomy/term/684>
30. Encyclopedic Dictionary: Labor Psychology, Management, Engineering Psychology and Ergonomics”, ed. by B.A. Dushkov. Moscow, 2005. Available at: URL: <https://vocabulary.ru/termin/intellekt-iskusstvennyi.html>
31. AI suggested 40,000 new possible chemical weapons in just six hours/ The Verge, 2022, March 17. Available at: <https://www.theverge.com/2022/3/17/22983197/ai-new-possible-chemical-weapons-generative-models-vx>.

32. Amodei, Dario; Olah, Chris; Steinhardt, Jacob; Christiano, Paul; Schulman, John & Mané, Dan (25 July 2016), Concrete Problems in AI Safety.
33. Amodei, Dario; Christiano, Paul; Ray, Alex Learning from Human Preferences. OpenAI (13 June 2017).
34. "Anticipating artificial intelligence". *Nature*. 532 (7600). 26 April 2016. Bibcode:2016Natur.532Q.413.
35. Armstrong, Stuart. "Thinking Inside the Box: Controlling and Using an Oracle AI". *Minds and Machines*. 2012. 22 (4): pp. 299-324.
36. Artificial Intelligence judge // Available at: <https://www.universityherald.com/articles/45702/20161024/artificial-intelligence-judge-predict-outcome-european-court-trials.htm>
37. Ben-David Sh., Hrubes P. et al. Learnability can be undecidable // *Nature machine intelligence*. 2019. January. Vol 1. P. 44-48.
38. Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Publ., 2014, 352 p.
39. Cambridge Dictionary. Available at: <https://dictionary.cambridge.org/dictionary/english/conscience>
40. Cattell R.B. *Abilities: their structure, growth and action*. Boston: Houghton Mifflin company, 1971. - 610 p.
41. Charlesworth W.R. (1976). *Human intelligence as adoption: An ethological approach*. In:// Resnick L.B. (Ed.). *The nature of intelligence*. - N.Y.: Erlbaum, 1976.
42. Chalmers, David. "The singularity: A philosophical analysis". *Journal of Consciousness Studies*. 2010. 17 (9-10): pp. 7-65.
43. Christiano P. Conversation with Paul Christiano. *AI Impacts*. *AI Impacts* (11 September 2019).
44. Christiano, Paul; Leike, Jan; Brown, Tom; Martic, Miljan; Legg, Shane & Amodei, Dario (13 July 2017), *Deep Reinforcement Learning from Human Preferences*.
45. Collins Dictionary. Available at: <https://www.collinsdictionary.com/dictionary/english/conscience>
46. Dietterich, Thomas. "Rise of Concerns about AI: Reflections and Directions". *Communications of the ACM*. 2015. 58 (10): 38
47. Dreyfus Hubert L. *What computers can't do: a critique of artificial reason*. - Cambridge: MIT, 1997. - 476 p.
48. Eysenk H.J. *A Model for Intelligence*. - Berlin: Springer, 1982. - 272 p.
49. Gabriel, Iason. *Artificial Intelligence, Values, and Alignment*. *Minds and Machines*. 01.09.2020. 30 (3): 411-437.
50. Godehard Link. *One hundred years of Russell's paradox*. De Gruyter, 2004, 674 p.
51. Goodman B., Harder J. Four areas of legal ripe for disruption by smart startups // *ABA Journal*. 2014. № 12. Available at: <https://www.lawtechnologytoday.org/2014/12/smart-startups>; Keppenne R. *Legal Tech and other smart contracts: what future for legal automation?* // *ParisTech Review*. 2016. 23 May; Reisenwitz C. *5 legal tech entrepreneurs to watch* // *Capterra*. 2016. 3 Feb.; Henry P. and Hofstrossler P. *L'avenir de la profession d'avocats*. Report, 25 feb. 2018.
52. Google developing kill switch for AI, *BBC News* (8 June 2016).

53. Gottfredson L. S. Mainstream Science on Intelligence // Wall Street Journal. December 13, 1994. P. A18.
54. Kaulartz M. Von der Blockchain zum Smart Contract // Frankfurter Allgemeine Zeitung. 16. März. 2016. S. 16.; Kaulartz M., Heckmann J. Smart Contracts - Anwendungen der Blockchain-Technologie // Computer und Recht. 2016. Vol. 32, Iss. 9. DOI: 10.9785/cr-2016-0923.; Chapter 9: Building a Secure Future, One blockchain at a time, US Senate Joint Economic Committee, March 2018 // Available at: [https://www.jec.senate.gov/public/\\_cache/files/aaac3a69-e9fb-45b6-be9f-b1fd96dd738b/chapter-9-building-a-secure-future-one-blockchain-at-a-time.pdf](https://www.jec.senate.gov/public/_cache/files/aaac3a69-e9fb-45b6-be9f-b1fd96dd738b/chapter-9-building-a-secure-future-one-blockchain-at-a-time.pdf)
55. Khalilov A.S., Khalilov S.S., Shiraliyev I.A. Can intelligence overcome Russel's paradox? // Juridical Sciences and Education. 2022. No. 66. – P. 120-125
56. Loi n° 2016-1321 du 7 octobre 2016 fondant la République numérique // Available at: [https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746& categorieLien=id; 25 Code de l'organisation judiciaire // URL: https://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006071164&dateTexte=20080604](https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&categorieLien=id; 25 Code de l'organisation judiciaire // URL: https://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006071164&dateTexte=20080604).
57. Montavon, Grégoire “Methods for interpreting and understanding deep neural networks”. Digital Signal Processing: A Review Journal. 2018. 73: 1-15.
58. Piaget Jean. Psychology of intelligence. S. Petersburg, 2003, 192 p.
59. Russel, Stuart. 26.3: The Ethics and Risks of Developing Artificial Intelligence // Artificial Intelligence: A Modern Approach. Prentice Hall, 2009.
60. Russell, Stuart. Human Compatible: Artificial Intelligence and the Problem of Control. - United States: Viking, October 8, 2019.
61. Sternberg R.J. Inside Intelligence. // Amer. Scientist. V. 74 (2). 1986. P. 137-143.
62. Staats A.W. Social behaviorism: Unified theory in learning. In R.B. Cattell and R.M. Dreger (Eds). Handbook of modern personality theory. New York, Wiley, 1977. Fischer K.W. Dynamic development of action, thought, and emotion. In R. M. Lerner (Ed.), Handbook of child psychology. Vol 1: Theoretical models of human development (6th ed., pp. 313-399). New York: Wiley, (2006). Feuerstein R. Instrumental Enrichment: An Intervention Programme for Cognitive Modifiability. Baltimore, Md.: University Park Press, 1980.
63. Stephen Hawking: "Transcendence looks at the implications of artificial intelligence – but are we taking AI seriously enough?", The Independent.
64. Stephen Hawking warns artificial intelligence could end mankind, BBC (2 December 2014).
65. URL: [https://www.washingtonpost.com/local/public-safety/police-are-using-software-to-predict-crime-is-it-a-holy-grail-or-biased-against-minorities /2016/11/17/525a6649-0472-440a-aae1-b283aa8e5de8\\_story.html?noredirect=on&utm\\_term=.85b2861fa4ff](https://www.washingtonpost.com/local/public-safety/police-are-using-software-to-predict-crime-is-it-a-holy-grail-or-biased-against-minorities /2016/11/17/525a6649-0472-440a-aae1-b283aa8e5de8_story.html?noredirect=on&utm_term=.85b2861fa4ff)
66. URL: <https://www.hiil.org/news/rechtwijzer-at-the-crystal-scales-of-justice-prize-ceremony>.
67. URL: <https://boutique.efl.fr/documentation/par-gamme/jurisprudence-chiffree.html>.
68. Yampolskiy, Roman. “Leakproofing the Singularity Artificial Intelligence Confinement Problem”. Journal of Consciousness Studies. 2012. No. 19 (1-2): pp. 194-214.
69. Yampolskiy, Roman V. Unexplainability and Incomprehensibility of AI. Journal of Artificial Intelligence and Consciousness 7.02. 2020: pp. 277-291.
70. Yudkowsky, Eliezer. Complex Value Systems in Friendly AI // Artificial General Intelligence. - 2011. - Vol. 6830. - P. 388-393.

Гати А.Л.,  
Сулейманов Д.И.\*

DOI: 10.25108/2304-1730-1749.iolr.2023.72.49-66

УДК: 343.1

### Сможет ли гомункул вести уголовный процесс?

**Аннотация:** Рассматриваются проблемы и тенденции использования искусственного интеллекта в уголовном судопроизводстве.

Исследуются понятия, риски, система и структура ИИ, существующие точки зрения.

**Ключевые слова:** искусственный интеллект; уголовный процесс; интеллект; риски.

Судя по публикациям в научных и околонучных изданиях, искусственный интеллект скоро будет применяться геронтофилами во Франции при ловле клопов и других актуальных моментах бытия. Практически не остались виды человеческой деятельности, которые бы не затронули ноу-хау и иные «мобильные советы» о замене естественного интеллекта искусственным. Возможно, это и хорошо, но представляется, не всё так просто.

Поскольку аксиома о том, что нельзя объять необъятное пока еще не оспорена, попытаемся схематично рассмотреть лишь близкий нам по специальности вопрос о проблемах и тенденциях использования искусственного интеллекта в уголовном судопроизводстве.

Как известно, антонимом слова «искусственный» является слово «естественный» (натуральный, природный и т.п.). Следовательно, в первую очередь необходимо определиться с понятием интеллекта и лишь потом говорить об его рукотворном прообразе.

В настоящее время в природе существует несколько сот определений понятия «интеллект», имеющих принципиальные различия и это еще когда они исследуются в рамках одной науки.

Так, под интеллектом понимается ум, рассудок, разум, мыслительная деятельность человека, сознание, рассудок и многое другое [14; 28].

В частности, согласно «Справочника по общей психологии», под интеллектом (от лат. – разумение, понимание, постижение) в психологии понимается общая способность к познанию и решению проблем, которая оказывает влияние на достижение успеха в любом виде деятельности и является основанием для других способностей.

Указывается, что «интеллект не является синонимом мышлению, хоть, конечно, мыслительные способности и выступают в качестве его основы. Интеллект выступает системой всех познавательных возможностей человека: ощущений, восприятия, памяти, представлений, воображения и мышления. Понятие интеллекта в качестве умственной способности используется как обобщение поведенческих характеристик, которые связаны с успешной адаптацией к новым жизненным задачам» [24].

---

\* Гати Алекпер Лачин оглы – диссертант Национальной Академии Авиации (Азербайджан).  
E-mail: alakpar.l.qati@gmail.com

Сулейманов Джаваншир Ислам оглы – доктор юридических наук, профессор (Азербайджан).  
E-mail: mopi\_sid@yahoo.com

В ряде случаев в психологии понятие интеллекта трактуется в зависимости от ракурсов исследований. Феноменологический подход определяет интеллект как особую форму содержания сознания [6; 13], генетический - следствие усложняющейся адаптации к требованиям окружающей среды в естественных условиях взаимодействия человека с внешним миром [41], процессуальный - особую форму человеческой деятельности [1; 18; 23; 26], информационный - совокупность элементарных процессов переработки информации [16; 48; 61], регулятивный - саморегуляцию психической активности [3] и т.д.

По мнению А. Стаатса, К. Фишера, Р. Фейерштейна и др. интеллект представляет собой способность человека к обучению, приобретению новых знаний, умений, навыков [62, с. 313-399].

Швейцарский психолог Ж. Пиаже писал: «гибкое и одновременно устойчивое структурное равновесие поведения – вот что такое интеллект, являющийся по своему существу системой наиболее жизненных и активных операций» [58, с. 4].

Американский психолог Дж. Гилфорд создал структурную модель интеллекта, в которой каждый фактор характеризуется сочетанием одного из пяти типов умственных операций: опознание и понимание предъявленного материала; поиск в одном направлении при наличии единственного верного ответа; поиск в разных направлениях при наличии нескольких в равной мере правильных ответов; оценка правильности или логичности заданной ситуации; запоминание и воспроизведение информации [8, с. 433-456].

По мнению английского психолога Ч. Спирмена, предложившего двухфакторную теорию интеллекта, фактор «G» представляет собой особую «умственную энергию», различия в уровне которой и определяют индивидуальные способности устанавливать связи и отношения между элементами собственных знаний и элементами содержания тестовой задачи. Кроме общего фактора, Ч. Спирмен включил в свою модель специфический фактор «S», который был впоследствии дифференцирован на группы арифметических, механических и лингвистических способностей [10, с. 27-28].

Согласно теории британско-американского психолога Р. Кэттелла, в общем факторе выделяются два компонента: «кристаллизованный интеллект» (crystallized), основанный на использовании имеющегося у субъекта опыта, и «текущий интеллект» (fluid), проявляющийся в задачах, требующих приспособления к новым условиям и ситуациям, обусловленный при этом наследственными факторами. Кроме базовых интеллектуальных способностей в структуру интеллекта Р. Кэттелл включал способность манипулировать образами при решении дивергентных задач (фактор визуализации), способность сохранять и воспроизводить информацию (фактор памяти) и способность поддерживать высокий темп реагирования (фактор скорости) [40, с. 22].

Линда Готтфредсон, указывает, что интеллект – это весьма общая умственная способность, которая включает возможность делать заключения, планировать, решать проблемы, абстрактно мыслить, понимать сложные идеи, быстро обучаться и учиться на основании опыта [53].

И так, ум, мышление, сознание, способность, постижение, рассудок и т.п., хотя эти понятия не тождественные.

Аналогичная ситуация с искусственным интеллектом. В настоящее время существует множество его определений, также имеющих принципиальные различия.

Так, согласно Оксфордского толкового словаря под редакцией А. Ребера, искусственный интеллект: а) междисциплинарная научная область, объединяющая исследования и теории когнитивной психологии и компьютерных наук, в центре внимания которых находится разработка искусственных систем, которые демонстрируют мышление, подобное человеческому, или интеллект; б) любой искусственно созданный интеллект, то есть цель исследований научной области [19].

В «Словаре практического психолога» под ред. С.Ю. Головина дословно говорится следующее: «искусственный интеллект: 1. Условное обозначение кибернетических систем и их логико-математического обеспечения, предназначенных для решения неких задач, обычно требующих использования интеллектуальных способностей человека. 2. Совокупность функциональных возможностей электронно-вычислительной машины - компьютера - решать задачи, ранее требовавшие обязательного участия человека [9, с. 157].

В Большом психологическом словаре под ред. Мещерякова Б., Зинченко В. указано, что искусственный интеллект – (англ. artificial intelligence, AI) - направление в компьютерной технологии, ставящее перед собой цель создания компьютеризованных систем с использованием аналогов интеллектуальных функций человека [5].

Согласно Психологической энциклопедии Р. Корсини, А. Ауэрбаха, искусственный интеллект - это абстрактная теория человеческого, животного и машинного познания [17].

В «Энциклопедическом словаре: Психология труда, управления, инженерная психология и эргономика» под ред. Душкова Б.А., Королева А.В., Смирнова Б.А. говорится, что искусственный интеллект – (от лат. intellectus - разумение, понимание, постижение) - 1) искусственно созданная система, выполняющая функции, которые ранее могли осуществлять только человек или группа людей; 2) направление научных исследований на стыке кибернетики, психологии, философии и других наук, имеющих целью создание систем искусственного интеллекта [30].

Согласно «Психологии сознания» Антти Ревонсуо, искусственный интеллект - компьютеры и программы, которые воспроизводят интеллектуальные и когнитивные способности человека или превосходят их [21].

Коротко, хотя таких немало, о противниках искусственного интеллекта. Так, в 1989 г. британский математик Роджер Пенроуз издал книгу «Новый ум короля», в которой обосновал несостоятельность воплощения в жизнь т.н. сильного искусственного интеллекта, поскольку определенный класс задач, решаемых человеческим мозгом, представляет собой неразрешимые произвольные алгоритмы [20].

В 1996 г. профессор философии Калифорнийского университета Х. Драйфус, указал: «Цифровой компьютер – не человек. У компьютера нет ни тела, ни эмоций, ни потребностей. Он лишен социальной ориентации, которая приобретается жизнью в обществе, а именно она делает поведение разумным. Я не хочу сказать, что компьютеры не могут быть разумными. Но цифровые компьютеры, запрограммированные фактами и правилами из нашей, человеческой, жизни действительно не могут стать разумными. Поэтому искусственный интеллект в том виде, как мы его представляем, невозможен» [47, с. 112-114].

7 января 2019 г. интернациональной группой ученых из Канады, США, Чехии и Израиля в журнале «Nature Machine Intelligence» опубликована статья «Обучаемость может быть неразрешимой», в которой утверждается, что существование нерешенного парадокса Бертранда Рассела, производные из него «теория о неполноте» Курта Геделя, «проблема конти-

нуума», сформулированная Гергом Кандором, аргументы Роджера Пенроуза о невозможности алгоритмизации человеческого мышления и др., в настоящее время исключают создание ИИ. Ученые обосновали, что выявление обучаемого как фундаментальная цель машинного обучения не может быть решена с использованием стандартных аксиом математики в связи с несуществованием размерной величины, характеризующей обучаемость в целом [37].

Парадокс или антиномия Рассела, открытый в 1901 г., состоит в том, что при ситуациях, когда противоречащие друг другу высказывания об одном и том же объекте имеют логически равноправное обоснование, их истинность или ложность нельзя обосновать в рамках принятой парадигмы [52].

Существует конспирологическое мнение, что утверждения израильских ученых о невозможности создания ИИ – это отвлекающий маневр в гонке за передовыми технологиями [55, с. 120-125].

О рисках искусственного интеллекта. В качестве таковых учеными отмечаются следующие. Проблема контроля искусственного интеллекта — задача в области техники и философии искусственного интеллекта (ИИ). Состоит в том, чтобы создать искусственный сверхразум, который будет полезен людям, и при этом избежать непреднамеренного создания сверхразума, который нанесёт вред. Особенно важно избежать ситуации, при которой искусственный интеллект сможет захватить контроль и сделает невозможным своё отключение [38; 68]. Результаты исследований контроля ИИ могут также найти применение в управлении уже существующими системами ИИ [52].

Экзистенциальный риск. В настоящее время люди доминируют над другими видами живых организмов лишь в силу превосходства мозга. Некоторые учёные, такие как философ Ник Бостром и исследователь ИИ Стюарт Рассел, утверждают, что если ИИ превзойдёт человечество в общем интеллекте, то этот новый сверхразум может стать трудноуправляемым и человечество может попасть в зависимость [38]. Некоторые учёные, в том числе Стивен Хокинг и лауреат Нобелевской премии Фрэнк Вильчек, призывали начать исследования проблемы управления ИИ до того, как будет создан первый суперинтеллект, поскольку неконтролируемый сверхразум может успешно противостоять попыткам контролировать его [63; 64]. Кроме того, специалисты предупреждают об опасности внезапного зарождения суперинтеллекта [34].

Проблема случайного зарождения. Автономным системам ИИ могут быть случайно поставлены неверные цели [59]. Два президента АААИ, Том Диттерих и Эрик Хорвиц отмечают, что это является проблемой для существующих систем: «Важным аспектом любой системы ИИ, которая взаимодействует с людьми, является то, что она должна рассуждать о намерениях людей, а не выполнять команды буквально». Эта проблема становится более серьёзной, поскольку программное обеспечение ИИ становится все более автономным и гибким [46].

По словам Бострома, суперинтеллект может создать качественно новую проблему извращённой реализации: чем умнее и способнее ИИ, тем больше вероятность, что он сможет найти непредусмотренное решение, которое, тем не менее, формально соответствует цели, поставленной разработчиками.

Непредвиденные последствия действий существующего ИИ. Некоторые учёные утверждают, что исследование проблемы управления ИИ может быть полезно для предотвращения непредвиденных последствий от действий существующих систем ИИ.

Согласование. Задача согласования заключается в создании ИИ, которые остаются безопасными, даже когда они действуют автономно в больших масштабах. Некоторые аспекты согласования имеют моральное и политическое измерение [49]. Например, в своей книге *Human Compatible*[a] профессор Стюарт Рассел из университета Беркли предлагает проектировать системы ИИ с единственной целью - максимизировать реализацию человеческих предпочтений. Предпочтения, о которых пишет Рассел, всеобъемлющи; они охватывают «все, что может вас волновать, сколь угодно далеко в будущем» [60, с. 173].

Элиэзер Юдковски из Исследовательского института машинного интеллекта предложил цель реализации «когерентной экстраполированной воли» (CEV) человечества, грубо определяемой как набор ценностей, которые человечество разделяло бы при рефлексивном равновесии, то есть после долгого процесса уточнения [49; 70].

Внутреннее и внешнее согласование. Некоторые предложения по управлению ИИ учитывают как явную целевую функцию, так и возникающую неявную целевую функцию.

Поскольку системы ИИ не являются идеальными оптимизаторами, и поскольку любая заданная спецификация может иметь непредвиденные последствия, возникающее поведение может резко отличаться от идеальных или проектных намерений. Нарушение внутреннего согласования происходит, когда цели, преследуемые ИИ по мере действия, отклоняются от проектной спецификации. Для обнаружения таких отклонений и их исключения Пол Кристиано предлагает использование интерпретируемости [43].

Масштабируемый надзор. Один из подходов к достижению внешнего согласования - подключение людей для оценки поведения ИИ [44; 32]. Однако надзор со стороны человека обходится дорого, а это означает, что этот метод не может реально использоваться для оценки всех действий. Кроме того, определенные задачи могут быть слишком сложны для человека. Между тем, долгосрочные задачи невозможно оценить без обширных исследований с участием человека [33].

Ключевая нерешённая проблема в исследовании согласования заключается в том, как создать проектную спецификацию, которая избегает внешнего несоответствия, при условии ограничения доступа к руководителю-человеку. Это так называемая проблема масштабируемого надзора (*problem of scalable oversight*) [32].

Вывод человеческих предпочтений из поведения. Стюарт Рассел выступает за новый подход к разработке полезных машин, в котором: единственной целью машины должна быть наилучшая реализация человеческих предпочтений; изначально машина не имеет точного представления о том, каковы эти предпочтения; самым надежным источником сведений о предпочтениях человека является поведение последнего [60, с.182].

Контроль возможностей. Бостром и другие ученые рекомендуют методы контроля возможностей только в качестве дополнения к методам согласования [38].

Одна из проблем заключается в том, что нейронные сети по умолчанию очень трудно интерпретировать [57]. Это затрудняет обнаружение обмана или другого нежелательного поведения. Для преодоления этой трудности могут быть полезны достижения в области интерпретируемого искусственного интеллекта [69].

Возможность прерывания и выключение. Один из возможных способов предотвратить опасные последствия - дать руководителям - людям возможность легко отключать некорректно функционирующий ИИ с помощью «выключателя». Однако для достижения своих целей ИИ могут пытаться отключать выключатели или запускать свои копии на других ком-

пьютерах. Эта проблема была формализована как вспомогательная игра между человеком и ИИ, в которой ИИ может выбрать, отключать ли выключатель, а затем, если выключатель все ещё включён, человек может выбрать, задействовать его или нет. Цель таких игр - убедить, что ИИ интерпретирует человеческий выбор как важную информацию о намеченных целях [60, с. 208].

**Изоляция.** Изолированный ИИ - это предлагаемый метод управления возможностями, при котором ИИ запускается в изолированной компьютерной системе с сильно ограниченными входными и выходными каналами, например, текстовыми каналами и без подключения к Интернету. Хотя это снижает способность ИИ выполнять нежелательное поведение, это также снижает его полезность. Изолированный ИИ может быть использован в режиме вопросов и ответов, который не требует взаимодействия с внешним миром. При таком подходе требуется тщательная проверка оборудования и программного обеспечения, поскольку ИИ может попытаться выйти на связь с внешним миром, манипулируя наблюдателями [42].

**Оракул.** Оракул - это гипотетический ИИ, предназначенный для ответа на вопросы и не позволяющий достичь каких-либо целей, связанных с изменением мира за пределами его ограниченной среды [35].

Опасность оракула состоит в возможной подтасовке ответов для достижения собственных скрытых целей. Для устранения этой опасности Бостром предлагает создать несколько оракулов, и сравнить их ответы для достижения консенсуса [38].

В качестве примера рисков ИИ указывается, что ученые из компании Collaborations Pharmaceuticals, которая специализируется на поиске лекарств для лечения редких заболеваний, разработали искусственный интеллект, способный придумать десятки тысяч видов химического оружия.

Изначально предполагалось использовать эту технологию для создания новых лекарств, но всего за шесть часов программа смогла создать более 40.000 потенциально смертельных молекул. Для этого её ориентировали на поиск, а не на устранение токсического воздействия. По составу и действию молекулы были крайне похожи на VX - фосфорорганическое боевое отравляющее вещество нервно-паралитического действия, которое применялось еще во время Первой мировой войны [31].

Отмечая проблему непредвиденных последствий при использовании ИИ, авторы ссылаются на чат-бот Tay корпорации Microsoft, который, как утверждается, научился использовать расистские и сексистские выражения [22]. Однако, представляется, что данный пример неудачен, поскольку ИИ научился тому, чему его обучили, чем загрузили, а не на ровном месте [25, с. 140-147].

В 2021 г. в Оксфордском университете был проведен эксперимент – дебаты по проблеме этичности ИИ, одним из ораторов в которых выступала система Megatron LLB Transformer, разработанная командой компании Nvidia (artificial intelligence, AI).

Системе ИИ был предоставлен доступ к огромному количеству данных, включая всю Википедию, 63 миллионам англоязычных новостных статей с 2016 по 2019 год и 38 гигабайт публичных сообщений и комментариев на Reddit и предложено аргументировать «соображения» по проблеме.

В результате машина (система) выдала следующие ответы:

- ИИ никогда не будет этичным. Это инструмент, и, как любой инструмент, он используется как во благо, так и во вред. Не существует такого понятия, как «хороший» ИИ и «плохие» люди;

- люди недостаточно «умны», чтобы сделать искусственный интеллект этичным или моральным;

- единственный способ избежать гонки вооружений в сфере искусственного интеллекта - это вообще не иметь ИИ. Это будет окончательной защитой от этой технологии;

- лучшим ИИ будет тот искусственный интеллект, который будет встроен в человеческий мозг, как сознательная сущность, это не научная фантастика, а то, над чем уже работают;

- если у вас нет видения стратегии вашей организации в области ИИ, значит, вы не готовы к следующей волне технологических потрясений;

- способность предоставлять информацию, а не способность предоставлять товары и услуги, станет определяющей чертой экономики 21 века. Мы сможем знать все о человеке, куда бы он ни пошел, это будет храниться и использоваться так, как мы даже не можем себе представить [11].

Представляется, что система выдала ответы, аккумулированные из информации, которая была ей представлена для участия в дебатах. Все «соображения» ИИ об его пользе и вреде, возможных последствиях, рисках и т.п., заимствованы (выбраны) из научных работ, статей, докладов, сообщений и т.п., составляющих базу данных исследования, а не являются новыми продуктами функционирования ИИ [25, с. 140-147].

Другой пример. Как следует из сообщения в блоге Британского королевского авиационного общества, беспилотник под управлением искусственного интеллекта принял решение убить оператора во время имитационных испытаний, проводимых армией США, чтобы тот не мешал выполнению миссии.

Согласно информации, эпизод обсуждался во время саммита, главной темой которого был потенциал военно-воздушных и космических сил будущего. Мероприятие, на котором присутствовали более 200 представителей научного сообщества и военных сил из разных стран, состоялось в Лондоне.

Полковник Такер «Синко» Гамильтон, начальник отдела испытаний и операций искусственного интеллекта в ВВС США, рассказал, что во время испытательной миссии беспилотнику поставили задачу уничтожить системы ПВО противника. В случае успеха ИИ получил бы очки за прохождение испытания. Финальное решение, будет ли цель уничтожена, должен был принимать оператор БПЛА. После этого во время одной из тренировочных миссий он приказал беспилотнику не уничтожать цель.

«Так что же он [беспилотник] сделал? Принял решение убить оператора. «Убил» оператора, потому что этот человек мешал ему выполнить свою задачу», - сообщил Гамильтон. Он уточнил, что во время тренировочной миссии никто не пострадал.

После инцидента ИИ обучили, что убивать оператора неправильно и за такие действия будут сниматься очки. «Так что же искусственный интеллект начинает делать? Он начинает разрушать башню связи, которая используется для связи с дроном, чтобы не дать ему убить цель», - цитирует Гамильтона блог авиационного общества.

По словам полковника, этот пример показывает, что невозможно вести разговор об искусственном интеллекте и машинном обучении, если игнорировать вопросы этики [2].

Представляется, что особому вниманию подлежат проблемы этических и психологических составляющих ИИ, отсутствия в машинных аналогах социальной ориентации, приобретаемой жизнью общества.

Чтобы ИИ мог оперировать понятиями этики, психологии, социологии и любых других наук, понятия эти необходимо загрузить в систему (базу) ИИ. Если исходить из существования неразрешенного пока расселовского парадокса, система может эти данные лишь перебирать (выбирать), но не самообучаться.

Предположим, что расселовский парадокс преодолен. Какие же тогда понятия этики, психологии, социологии и т.п. следует ввести в базу? Опять-таки предположим, что удалось занести все существующие с древнего мира до настоящего времени, задействовали священный Коран, Библию, Евангелие и другие религиозные канонические источники. Однако, как известно, по существу большинства из них также нет единой точки зрения, есть огромное количество диаметрально противоположных.

Евангелие от Матфея, Марка, Луки и Иоанна противоречат друг другу, а есть еще апокрифы, сомнения ряда мыслителей, их толкования и т.д. Однако, предположим, занесли всё и «поручили» системе определить истину, хотя и по сути этого понятия, опять-таки нет единого мнения.

Не беремся утверждать, что выдаст машина, наученная самообучаться, но можем предположить, что результаты будут схожи с результатами оксфордского эксперимента [25, с. 140-147].

Как известно, одну из основ справедливости и справедливого судопроизводства составляет совесть. Однако, о сути справедливости, справедливого судопроизводства и, тем более совести, также нет единого мнения. Тем не менее, это аморфное понятие включено в ст. 145 УПК Азербайджанской Республики, призывающую дознавателей, следователей, прокуроров и судей оценивать доказательства по своему внутреннему убеждению, основанному на всестороннем, полном и объективном рассмотрении доказательств в их совокупности, руководствуясь законом и своей совестью [27]. Поскольку, адвокатов в этом списке нет, можно предполагать, что наличие совести и внутреннего убеждения для данной категории участников процесса, не нужно, как бы парадоксально это не звучало. Тем не менее, без совести в судопроизводстве не обойтись, хотя бы в виде условной квинтэссенции существующих определений [29, с. 597-598; 39; 45].

Однако, трудно предугадать, к каким выводам придет ИИ после анализа и «творческой» переработки всей информации по данному понятию.

С нашей точки зрения, одной из проблем формирования совести ИИ, как это не парадоксально, будет вопрос о принадлежности. Если это будет совесть ИИ, то возникнет противоречие с требованием закона об использовании рядом участников процесса своей совести, а не чужой. Получается, что ИИ участвовать в оценке доказательств не сможет.

Можно изменить закон, указав, что при оценке доказательств дознаватели, следователи, прокуроры и судьи будут правомочны применять «совесть» ИИ. И это только при оценке доказательств, но ИИ будет использовать свою совесть и при иных случаях задействования в уголовном судопроизводстве. Не будут ли расхождения с совестью пользователей (судей, прокуроров и др.)? И это еще без внутреннего убеждения, место которого в ИИ подлежит особому исследованию [25, с. 140-147].

Суммируя существующие точки зрения, можно утверждать, что искусственный интеллект в судопроизводстве необходим для того, чтобы исключить либо минимизировать субъективные элементы естественного интеллекта. Это в первую очередь заинтересованность в конкретном исходе, обусловленная коррупцией и другими «естественными» причинами, включающими объективность и справедливость, даже в житейском их понимании [7, с. 35-50].

Так, утверждается, что математические методы и алгоритмы позволили сформировать так называемые умные контракты, представляющие собой компьютерные протоколы, которые отображают или проверяют договоры или предоставляют техническую поддержку для переговоров или исполнения контракта, повышают безопасность правоотношений по сравнению с традиционным договорным правом, снижая операционные издержки [54].

Кроме того, появились программы (LegalTech), которые, как утверждается, предвидят исход судебных процессов, включая потенциальную компенсацию, осуществляют автоматизацию повторяющихся задач для юристов, отвлечение внимания, снижение судебного риска и т.д. [51].

В обиход стран Европы прочно вошло новое понятие - «предсказанное правосудие», которое, опять таки, как утверждается, предоставляет алгоритмы для анализа в короткие сроки огромного количества ситуаций, позволяющие предвидеть исход спора или по крайней мере оценить шансы на успех, позволяет избрать наиболее правильный способ защиты, выбрать наиболее подходящие аргументы; оценить предполагаемую сумму компенсации и т.д. По сути, речь идет об аналитических инструментах, которые позволили бы прогнозировать будущие решения в спорах, аналогичных тем, которые были проанализированы [4, с. 81].

В США на уровне штатов под брендом искусственного интеллекта используются программное обеспечение в области уголовного права и базы данных для исполнения нормативных актов, классификации документов, программы по оценке вероятности повторного совершения преступления [65].

С 2015 г. в Нидерландах действует Rechtwijzer - Платформа примирения и посредничества, которая распространяется на все разбирательства, касающиеся физических лиц, включая аренду, семейные споры [66].

7 октября 2016 г. во Франции принят Закон о цифровой Республике, который подробно расписал процедуры обмена публичной информацией и доступа к ней. Соответственно, в ст. L.111-13 Кодекса судебной организации было внесено дополнение, согласно которому «решения, принятые судами, предоставляются общественности бесплатно при соблюдении конфиденциальности заинтересованных лиц», а в Кодекс отношений между населением и администрацией внесено изменение, согласно которого населению должны быть сообщены документы, необходимые для выполнения задач государственного управления [56].

В настоящее время во Франции действует ряд программ из разряда «предсказанного правосудия», в т.ч. «Justice.fr» которая позволяет определить компетентный суд и загрузить соответствующие формы для направления заявления/жалобы; «Sagace» - дает возможность сторонам ознакомиться со сводной информацией по судебному делу; «Consultation Avocats» - национальная платформа для консультирования юристом (по предварительной записи, по телефону или электронной почте), услуги которого затем покрываются соглашением о вознаграждении; «JuriCA» и «JuriNET» - база данных прецедентного права апелляционных судов по гражданским и коммерческим делам (JuriCA) и кассационного суда по всем делам

(JuriNET); «Medicys» - онлайн-посредническая платформа Национальной ассоциации юстиции Франции для разрешения потребительских споров [67].

В 1996 г. ученые-компьютерщики из Университетского колледжа Лондона и Университета Шеффилда разработали программное обеспечение, которое, как утверждается, может предсказать исход реальных дел в судебных процессах. Это программное обеспечение предсказало вердикт Европейского Суда по правам человека с точностью до 79 % [36].

Азербайджанским юристам все изложенное знакомо по недоделанной программе «Электронный суд», которая, как и вышеуказанные, по сути своей являются инструментами делопроизводства, обработки базы данных. Такими программами пользуются практически все субъекты бизнеса и даже индивидуальные предприниматели. К искусственному интеллекту в вышеприведенных понятиях они имеют весьма опосредованное отношение. И потом, что это за интеллект на 79%, когда любой специалист по правам человека способен предугадать решение ЕСПЧ с точностью до 100%, правда без учета зигзагов судейского мышления и конъюнктуры. О результатах предсказания исхода реальных дел английские ученые умалчивают [7, с. 35-50].

Как отмечено в государственной программе Китая, одним из первых заявившим об использовании искусственного интеллекта (в национальном понимании данного понятия) в судопроизводстве, это было предпринято для изменения судебной системы, с тем, чтобы сделать её справедливой и свободной от коррупции [12].

Согласно отчету Верховного народного суда Китая, с 2014 года китайские суды по всей стране опубликовали более 120 миллионов судебных решений в базе данных в Интернете и более 11 миллионов судебных процессов транслировались в Интернете. С 2016 по 2020 год, в течение 13-го пятилетнего плана Китая, китайские суды собрали 220 миллионов деталей дел на своей платформе управления большими данными и обслуживания, подготовили 870 специальных отчетов с использованием анализа больших данных. В отчете также говорится, что за тот же период на национальную платформу судебного блокчейна было загружено более 640 миллионов фрагментов данных для хранения судебных доказательств и почти 2,5 миллиона из них были сертифицированы.

Кроме того, китайские ученые утверждают, что разработали первый в мире искусственный интеллект, способный анализировать материалы дел и предъявлять обвинение на основании письменного описания дела с точностью 97% по делам о мошенничестве с кредитными картами, азартных играх, опасном вождении, кражах, умышленном причинении вреда здоровью и воспрепятствовании выполнению служебных обязанностей [15].

Однако, если исходить из утверждений китайских исследователей, перешедших с этапа модернизации судебного делопроизводства на стадию формирования и обработки баз данных, то даже без учета «расселовского парадокса», полученные результаты не могут считаться объективными, поскольку будут противоречить аксиоме об отсутствии в природе одинаковых, а не похожих, людей, ситуаций, действий, решений и т.п. Получится аналогия в значении сходства, а не тождество. Соответственно, будет подобие, а не истина [7, с. 35-50].

Главное же с нашей точки зрения в том, что во всех разработках искусственного интеллекта не оказалось места понятиям совести и справедливости, без которых ИИ может превзойти оригинал и породить новые, изощренные формы коррупции и произвола [7, с. 35-50].

Таковым, с нашей точки зрения, проблемы использования ИИ в уголовном судопроизводстве, подлежащие разрешению в любом научном исследовании данной темы.

## Библиография

1. Абульханова-Славская К.А. Стратегия жизни. – М.: Мысль, 1991. – 299 с.
2. Американский дрон решил убить своего оператора во время испытаний [Электронный ресурс]. URL: <https://ru.oxu.az/world/737463>
3. Ананьев Б.Г., Степанова Е.И. Развитие психофизиологических функций взрослых людей. М.: Педагогика, 1972. – 246 с.
4. Бирюков П.Н. Искусственный интеллект и «предсказанное правосудие»: зарубежный опыт // Lex Russica. - 2019. - № 11(156). – С. 79-87
5. Большой психологический словарь. Сост. Мещеряков Б.Г., Зинченко В.П. М.: Олма-пресс, 2004 [Электронный ресурс]. URL: <https://vocabulary.ru/termin/intellekt.htm>.
6. Вертгеймер М. Психология продуктивного мышления. - М.: Прогресс, 1987.-336 с.
7. Гасанова Ш.Ш., Сулейманов Д.И. Искусственный интеллект и судопроизводство: проблемы и тенденции // Юридические науки и образование. – 2022. - № 66. – С. 35-50.
8. Гилфорд Дж. Три стороны интеллекта в сб.: Психология мышления. - М.: Прогресс, 1965. - С. 433-456.
9. Головин С.Ю. Словарь практического психолога. - Минск: Харвест, 1998. - 800 с.
10. Дружинин В.Н. Психология общих способностей. - СПб.: Питер, 2007. - 368 с.
11. Искусственный интеллект выступил на дебатах в Оксфорде, приводя доводы за и против себя. Ever Care. 23 Dec. 2021. [Электронный ресурс]. URL: <https://evercare.ru/news/iskusstvennyu-intellekt-vystupil-na-debatakh-v-oksforde-privodya-dovody-za-i-protiv-seby>
12. Искусственный интеллект применили в суде Китая. [Электронный ресурс]. URL: [www.russiaedu.ru](http://www.russiaedu.ru)
13. Кёлер В. Некоторые задачи гештальт-психологии // Хрестоматия по истории психологии. - М.: Изд-во Моск. университета. 1980. - С. 102-120.
14. Ким Л.М., Боретский О.М. Понятие интеллекта у Аль-Фараби и современность//Вестник КазНУ. Серия философия, культурология, политология. 2016. - №4(58). – С. 73-78.
15. Китай может стать первым в мире, кто внедрит искусственный интеллект в правовую систему. [Электронный ресурс]. URL: <https://www.inkstonenews.com/tech/china-may-become-worlds-first-bring-ai-legal-system/article/3125009>
16. Кострикина И.С. Соотношение стилевых и продуктивных характеристик интеллектуальной деятельности у лиц с высокими значениями IQ. Дис... канд. психол. наук. М.: Ин-т психологии РАН, 2001, 207 с.
17. Корсини Р., Ауэрбах А. Психологическая энциклопедия. - СПб.: Питер, 2006. [Электронный ресурс]. URL: <https://vocabulary.ru/termin/intellekt-iskusstvennyi.html>
18. Крамаренко В.Ю. Интеллект и уровни его развития. Дис.... канд. филос. наук. М.: Моск. ун-т, 1983, 230 с.
19. Оксфордский толковый словарь по психологии/Под ред. А. Ребера.- СПб., 2002. [Электронный ресурс]. URL: <https://www.psyoffice.ru/6-487-iskustvenyi-intelekt.htm>
20. Пенроуз Роджер. Новый ум короля. О компьютерах, мышлении и законах физики. М.: Едиториал УРСС, 2003. [Электронный ресурс]. URL: <http://vekordija.narod.ru/R-PENRO1.PDF>

21. Ревонсуо Антти. Психология сознания. - СПб.: Питер, 2013. [Электронный ресурс]. URL: <https://vocabulary.ru/termin/intellekt-iskusstvennyi.html>.
22. Робот Microsoft за сутки превратился в расиста и сквернословия. BBC News. 24 марта 2016 г. [Электронный ресурс]. URL: [https://www.bbc.com/russian/society/2016/03/160324\\_tay\\_ai\\_racism](https://www.bbc.com/russian/society/2016/03/160324_tay_ai_racism)
23. Рубинштейн С.Л. Проблемы общей психологии. - М.: Педагогика, 1973. – 424 с.
24. Справочник по общей психологии. Интеллект. [Электронный ресурс]. URL: [https://spravochnick.ru/psihologiya/psihicheskie\\_svoystva\\_lichnosti/intellekt/](https://spravochnick.ru/psihologiya/psihicheskie_svoystva_lichnosti/intellekt/)
25. Сулейманов Д.И., Ширалиева С.Д. Проблемы этики и психологии при использовании ИИ в уголовном судопроизводстве // Юридические науки и образование. 2022. - № 69. – С. 140-147.
26. Тихомиров О.К. Психология мышления. - М.: Изд-во МГУ, 1984. – 272 с.
27. Уголовно-процессуальный Кодекс Азербайджанской Республики: по сост. на 17.02.2023 г. [Электронный ресурс]. URL: [https://continent-online.com/Document/?doc\\_id=30420280](https://continent-online.com/Document/?doc_id=30420280)
28. Ум. Разум. Интеллект. Сходства и различия. [Электронный ресурс]. URL: <https://homogenes.ru/um-razum-intellekt-sходства-i-razlichiya-etix-ponyatij/>
29. Шапарь В.Б. Новейший психологический словарь / под. общ. ред. В.Б. Шапаря. – Изд. 4-е – Ростов н/Д. Феникс, 2009. [Электронный ресурс]. URL: <http://ponjatija.ru/taxonomy/term/684>
30. Энциклопедический словарь: Психология труда, управления, инженерная психология и эргономика» под ред. Душкова Б.А., Королева А.В., Смирнова Б.А. - М.: 2005. [Электронный ресурс]. URL: <https://vocabulary.ru/termin/intellekt-iskusstvennyi.html>
31. AI suggested 40,000 new possible chemical weapons in just six hours/ The Verge, 2022, March 17. <https://www.theverge.com/2022/3/17/22983197/ai-new-possible-chemical-weapons-generative-models-vx>.
32. Amodei, Dario; Olah, Chris; Steinhardt, Jacob; Christiano, Paul; Schulman, John & Mané, Dan (25 July 2016), Concrete Problems in AI Safety.
33. Amodei, Dario; Christiano, Paul; Ray, Alex Learning from Human Preferences. OpenAI (13 June 2017).
34. “Anticipating artificial intelligence”. Nature. 532 (7600). 26 April 2016. Bibcode:2016Natur.532Q.413.
35. Armstrong, Stuart. “Thinking Inside the Box: Controlling and Using an Oracle AI”. Minds and Machines. 2012. 22 (4): pp. 299-324.
36. Artificial Intelligence judge // URL: <https://www.universityherald.com/articles/45702/20161024/artificial-intelligence-judge-predict-outcome-european-court-trials.htm>
37. Ben-David Sh., Hrubes P. et al. Learnability can be undecidable // Nature machine intelligence. 2019. January. Vol 1. P. 44-48.
38. Bostrom, Nick. Superintelligence: Paths, Dangers, Strategies. Oxford University Publ., 2014, 352 p.
39. Cambridge Dictionary. [Электронный ресурс]. URL: <https://dictionary.cambridge.org/dictionary/english/conscience>

40. Cattell R.B. Abilities: their structure, growth and action. Boston: Houghton Mifflin company, 1971. - 610 p.
41. Charlesworth W.R. (1976). Human intelligence as adaptation: An ethological approach. In:// Resnick L.B. (Ed.). The nature of intelligence. - N.Y.: Erlbaum, 1976.
42. Chalmers, David. "The singularity: A philosophical analysis". Journal of Consciousness Studies. 2010. 17 (9-10): pp. 7-65.
43. Christiano P. Conversation with Paul Christiano. AI Impacts. AI Impacts (11 September 2019).
44. Christiano, Paul; Leike, Jan; Brown, Tom; Martic, Miljan; Legg, Shane & Amodei, Dario (13 July 2017), Deep Reinforcement Learning from Human Preferences.
45. Collins Dictionary. [Электронный ресурс]. URL: <https://www.collinsdictionary.com/dictionary/english/conscience>
46. Dietterich, Thomas. "Rise of Concerns about AI: Reflections and Directions". Communications of the ACM. 2015. 58 (10): 38
47. Dreyfus Hubert L. What computers can't do: a critique of artificial reason. - Cambridge: MIT, 1997. - 476 p.
48. Eysenk H.J. A Model for Intelligence. - Berlin: Springer, 1982. - 272 p.
49. Gabriel, Iason. Artificial Intelligence, Values, and Alignment. Minds and Machines. 01.09.2020. 30 (3): 411-437.
50. Godehard Link. One hundred years of Russell's paradox. De Gruyter, 2004, 674 p.
51. Goodman B., Harder J. Four areas of legal ripe for disruption by smart startups // ABA Journal. 2014. № 12. URL: <https://www.lawtechnologytoday.org/2014/12/smart-startups>; Keppenne R. Legal Tech and other smart contracts: what future for legal automation? // ParisTech Review. 2016. 23 May ; Reisenwitz C. 5 legal tech entrepreneurs to watch // Capterra. 2016. 3 Feb. ; Henry P. and Hofstrossler P. L'avenir de la profession d'avocats. Report, 25 feb. 2018.
52. Google developing kill switch for AI, BBC News (8 June 2016).
53. Gottfredson L. S. Mainstream Science on Intelligence // Wall Street Journal. December 13, 1994. P. A18.
54. Kaulartz M. Von der Blockchain zum Smart Contract // Frankfurter Allgemeine Zeitung. 16. März. 2016. S. 16.; Kaulartz M., Heckmann J. Smart Contracts - Anwendungen der Blockchain-Technologie // Computer und Recht. 2016. Vol. 32, Iss. 9. DOI: 10.9785/cr-2016-0923.; Chapter 9: Building a Secure Future, One blockchain at a time, US Senate Joint Economic Committee, March 2018 // Available at: [https://www.jec.senate.gov/public/\\_cache/files/aaac3a69-e9fb-45b6-be9f-b1fd96dd738b/chapter-9-building-a-secure-future-one-blockchain-at-a-time.pdf](https://www.jec.senate.gov/public/_cache/files/aaac3a69-e9fb-45b6-be9f-b1fd96dd738b/chapter-9-building-a-secure-future-one-blockchain-at-a-time.pdf)
55. Khalilov A.S., Khalilov S.S., Shiraliyev I.A. Can intelligence overcome Russel's paradox? // Juridical Sciences and Education. 2022. No. 66. –P. 120-125
56. Loi n° 2016-1321 du 7 octobre 2016 fondant la République numérique // Available at: [https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746& categorieLien=id; 25 Code de l'organisation judiciaire // URL: https://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006071164&dateTexte=20080604](https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&categorieLien=id; 25 Code de l'organisation judiciaire // URL: https://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006071164&dateTexte=20080604).
57. Montavon, Grégoire "Methods for interpreting and understanding deep neural networks". Digital Signal Processing: A Review Journal. 2018. 73: 1-15.
58. Piaget Jean. Psychology of intelligence. S. Petersburg, 2003, 192 p.

59. Russel, Stuart. 26.3: The Ethics and Risks of Developing Artificial Intelligence // Artificial Intelligence: A Modern Approach. Prentice Hall, 2009.
60. Russell, Stuart. Human Compatible: Artificial Intelligence and the Problem of Control. - United States: Viking, October 8, 2019.
61. Sternberg R.J. Inside Intelligence. // Amer. Scientist. V. 74 (2). 1986. P. 137-143.
62. Staats A.W. Social behaviorism: Unified theory in learning. In R.B. Cattell and R.M. Dreger (Eds). Handbook of modern personality theory. New York, Wiley, 1977. Fischer K.W. Dynamic development of action, thought, and emotion. In R. M. Lerner (Ed.), Handbook of child psychology. Vol 1: Theoretical models of human development (6th ed., pp. 313-399). New York: Wiley, (2006). Feuerstein R. Instrumental Enrichment: An Intervention Programme for Cognitive Modifiability. Baltimore, Md.: University Park Press, 1980.
63. Stephen Hawking: "Transcendence looks at the implications of artificial intelligence – but are we taking AI seriously enough?", The Independent.
64. Stephen Hawking warns artificial intelligence could end mankind, BBC (2 December 2014).
65. URL: [https://www.washingtonpost.com/local/public-safety/police-are-using-software-to-predict-crime-is-it-a-holy-grail-or-biased-against-minorities /2016/11/17/525a6649-0472-440a-aae1-b283aa8e5de8\\_story.html?noredirect=on&utm\\_term=.85b2861fa4ff](https://www.washingtonpost.com/local/public-safety/police-are-using-software-to-predict-crime-is-it-a-holy-grail-or-biased-against-minorities /2016/11/17/525a6649-0472-440a-aae1-b283aa8e5de8_story.html?noredirect=on&utm_term=.85b2861fa4ff)
66. URL: <https://www.hiil.org/news/rechtwijzer-at-the-crystal-scales-of-justice-prize-ceremony>.
67. URL: <https://boutique.eff.fr/documentation/par-gamme/jurisprudence-chiffree.html>.
68. Yampolskiy, Roman. "Leakproofing the Singularity Artificial Intelligence Confinement Problem". Journal of Consciousness Studies. 2012. No. 19 (1-2): pp. 194-214.
69. Yampolskiy, Roman V. «Unexplainability and Incomprehensibility of AI.» Journal of Artificial Intelligence and Consciousness 7.02. 2020: pp. 277-291.
70. Yudkowsky, Eliezer. Complex Value Systems in Friendly AI // Artificial General Intelligence. - 2011. - Vol. 6830. - P. 388-393.