

УДК 330.46

Реальные и мнимые цели искусственного интеллекта

О. Л. Фиговский¹, О. Г. Пенский²

¹Israel, 10550, Migdal, HaEmek, 73

figovsky@gmail.com

²Пермский государственный национальный исследовательский университет

Россия, 614900, г. Пермь, ул. Букирева, 15

ogpensky@mail.ru; +7 342 239 63 09

Описываются и обосновываются возможные опасности искусственного интеллекта для психологии человека, приводятся примеры проявления этой опасности в современном мире, выдвигается и обосновывается гипотеза о том, что, благодаря влиянию искусственного интеллекта на человека, возможны изменения в способах мышления человека, предлагается математическая модель, позволяющая вычислять влияние искусственного интеллекта на психологические параметры человека, предлагается для контроля влияния искусственного интеллекта на социум сформулировать конкретные цели внедрения искусственного интеллекта в общество, учитывающие отрицательное влияние этого интеллекта на психологию человека. На основе заданной цели внедрения искусственного интеллекта в социум предлагается несложная математическая модель, позволяющая проводить численную экспресс-оценку влияния общества на "психологию" робота и наоборот. Модель иллюстрируется простым примером вычисления этого влияния в современном обществе в целом.

Ключевые слова: *робот; искусственный интеллект; общество; социум; психология человека; психология робота; математические модели психологии робота.*

DOI: 10.17072/1993-0550-2021-1-43-48

В настоящее время все развитые государства мира устремились в безудержную и беспрецедентную в истории гонку за искусственным интеллектом, его внедрением, а, порой, жестким проталкиванием в жизнь общества.

Однако до сих пор, насколько известно авторам настоящей статьи, нет единого определения искусственного интеллекта (ИИ), а поэтому до конца не понятно:

– Что же именно внедряют в социум политики и бизнесмены?

Мы не будем анализировать или критиковать эти многочисленные определения, а попробуем дать свое определение: "Искусственный интеллект – это автоматизированное мышление, способное самостоятельно принимать решения".

Анализируя устремления политиков и ученых многих стран, мы можем сделать вывод о том, что они, порой, сами того не понимая, повсеместно внедряя методы искусственного интеллекта, преследуют именно ту цель, которая прописана в приведенном выше определении: "Заставить автоматы принимать решения за живых людей", тем самым дать возможность машинам определять многие человеческие судьбы.

Примером реализации такой возможности является КНР. Пока российские суды осваивают онлайн-заседания, видео-конференц-связь и голосование через блокчейн, китайская судебная система ушла далеко вперед. В судах Китая роботы уже помогают судьям принимать решения, составляют сторонам процессуальные документы и ведут печатные онлайн-трансляции заседаний без помощи человека.

Причем такие технологии применяются по всей стране в судах разного уровня. На развитие и внедрение новшеств у китайцев ушло примерно пять лет [1]. Искусственный интеллект помогает судьям в рассмотрении простых дел, например, споров по контрактам с низкой стоимостью. Таким образом, судебные органы страны пытаются достичь единообразия в практике по одинаковым делам [1]. Искусственный интеллект используется в китайских судах с начала 2019 года, пока это использование ограничивается главным образом представлением доказательств, связанных с делом, и помощью в расследованиях [2], т.е. искусственный интеллект уже принимает решения о том, включать или не включать имеющиеся у следствия материалы в качестве доказательств преступлений, совершенных человеком.

Конечно, приведенный пример жесткого использования в социуме искусственного интеллекта является пока исключением из правил, но он очень красочно говорит о реальных целях разработчиков алгоритмов этого интеллекта, которые на сегодня самими разработчиками, может быть, пока не осознаются, но которые подтверждают приведенное нами в начале статьи определение.

Существует несколько типов искусственного интеллекта [3]. Приведем определение одного из них, которое сформулировано в 2018 г.: "Автономный ИИ предполагает возможность системы функционировать долгое время без участия оператора. Например, он позволяет дрону, оборудованному солнечными батареями, совершить многодневное путешествие с Елисейских полей на Красную площадь или в обратном направлении, самостоятельно выбирая как маршрут, так и места для промежуточных посадок для подзарядки аккумуляторов, избегая при этом всевозможные препятствия" [3]. Мы видим, что в примере для описания автономного ИИ нет ничего опасного, но определение автономного ИИ говорит о том, что в настоящее время создаются приборы, способные работать без вмешательства и контроля со стороны человека – а это уже опасно.

Сейчас наука развивается гораздо быстрее, чем осознание разумом человека (пускай, даже самого умного) всех последствий ускоренного развития этой науки.

Алгоритмы и аппарат искусственного интеллекта создал сам человек. Но создавал

он этот аппарат не "по образу и подобию своему", а на основе собственных гипотез о том, как думает и устроен сам, и называя порой похожие на человека лишь внешне и схематически структуры искусственного интеллекта. Именно так родились очень популярные сейчас схемы, названные *нейросетевыми* [4], которые на самом деле имеют не очень много общего с реальными нейронами человеческого мозга. Но именно искусственные (не природные) нейросети открыли большие возможности для сверхбыстрого решения многих задач и приобрели огромную популярность во всем мире.

Экспертные системы [5], применяемые для получения новых результатов при обработке, например, больших массивов данных, тоже создал человек, но и эти системы, как развитая часть методов искусственного интеллекта, сами являются искусственными и лишь немного похожими на некоторые способности мышления человека.

Аналогичными особенностями обладают и другие направления искусственного интеллекта.

Все, что создано человеком в сфере искусственного интеллекта является по самой своей сути искусственными для человека конструкциями и используются лишь аналоги, дополненные неким вымыслом, связанным порой с фантастическим искусством математики. Поэтому человек создал интеллект по сути своей отличающийся от психологии человека, а, следовательно, чуждый его природе. Но, благодаря самому человеку, быстро развивающийся искусственный интеллект требует от человека искать безопасные пути сосуществования с чуждым по своей природе интеллектом.

В настоящее время порожденный человеком и находящийся пока в начале своего исторического развития искусственный разум начинает серьезно влиять на психологию самого создателя.

Этот эффект, на наш взгляд, может наиболее сильно повлиять на молодое поколение. Ни для кого уже не является секретом, что сейчас роботы активно внедряются в сферу образования. Учителя школ, преподаватели вузов, пусть пока не во всех сферах их деятельности, но уже начинают заменяться искусственным интеллектом. Поэтому возникает риторический вопрос:

– А не станут ли ученики и студенты мыслить алгоритмами и схемами искусствен-

ного интеллекта, потеряв при этом свой природный разум?

Приведем небольшой пример влияния на аудиторию средств массовой информации для формирования общественного сознания аудитории в нужном кому-либо направлении.

В работе [6] впервые введен и подробно описан термин "социальные нанотехнологии", которому можно дать следующее определение: "Социальные нанотехнологии – это невербальное (на уровне подсознания) влияние субъекта на аудиторию". Это влияние может оказываться многими способами, даже, например, подбором ведущих телевизионных или интернет-программ с необходимым заказчику внешним видом, темпераментом, уровнем общей культуры, тембром речи и т.д. Но уже сейчас телеведущие начали заменяться роботами, т.е. искусственным интеллектом. Об этом довольно много написано, например, в доступном всем Интернете [7, 8]. Замена человека роботом-ведущим обусловлена малыми экономическими затратами на его круглосуточное функционирование в эфире. В настоящее время стали широко известны роботы-художники, роботы-музыканты, роботы-композиторы, роботы-актеры и т.д. Эти роботы – творческие личности своей "профессиональной" деятельностью призваны влиять на подсознание живых людей, становясь автоматическими "социальными нанотехнологиями", воздействующими своим "неодушевленным" искусством на духовную сферу жизни человека.

– Кажется бы, что может быть плохого во влиянии искусственного искусства на человека?

– Но робототехнический искусственный интеллект сейчас стоит дорого и в развитие его алгоритмов вносят деньги очень состоятельные люди и политики, которые по своей сути являются прагматиками. Поэтому неизбежно проявление психологии прагматизма и у искусственного искусства, которое породит прагматизм у живых людей. А это все и является робототехническими социальными нанотехнологиями.

В монографиях [9, 10] описываются математические модели цифровых двойников человека. Пока эти модели основаны на математизации "бытовой" психологии отдельного субъекта и групп субъектов. Эта математизация, в числе прочего, позволила объективно описывать пока простейшее поведение систем "робот–человек".

На основе математических моделей была доказана теорема о том, что робот с абсолютной памятью (робот, который в отличие от человека помнит все) опасен для человека [11], где под опасностью следует понимать психологическое давление более воспитанного (здесь воспитание – это одновременно и психологический, и математический термин [9, 10]) по абсолютной величине робота на менее воспитанного по абсолютной величине человека.

В работах [12, 13] впервые введены так называемые *коэффициенты мягкого и жесткого влияния одного субъекта на другого*, а разработанная компьютерная программа [14] позволяет численно измерять эти коэффициенты.

В связи с этим удалось сформулировать следующее утверждение: "Влияние искусственного интеллекта на воспитание человека с целью полного принятия человеком убеждений робота неизбежно". Доказательство утверждения основано на следующих очевидных фактах: "При общении робота с человеком коэффициент психологического влияния человека на бездушного робота равен нулю, а коэффициент влияния робота на одухотворенного человека отличен от нуля. Поэтому робот влияет на воспитание человека, а не наоборот".

– Как выходить из этой непростой ситуации?

Ответ на этот вопрос может быть таким:

– Для психологической безопасности человека необходимо создавать эмоциональных роботов с коэффициентами влияния на человека, меньшими, чем коэффициенты влияния человека на роботов.

Недавно в СМИ была опубликована статья с пугающим красноречивым названием "Три главных угрозы существованию человечества" [15]. Приведем ниже выдержку из этой статьи.

"Искусственный интеллект, синтетическая биология и так называемые неизвестные неизвестные могут уничтожить человечество до 2100 года", считает сооснователь Skype Ян Таллинн. Из трех угроз, которые больше всего беспокоят Таллинна, он сосредоточен на искусственном интеллекте и тратит миллионы долларов, пытаясь обеспечить безопасное развитие технологии. Это включает в себя инвестиции на ранних стадиях в лаборатории искусственного интеллекта, такие как DeepMind (отчасти для того, чтобы он мог

следить за тем, что они делают) и финансирование исследований безопасности ИИ в таких университетах, как Оксфорд и Кембридж.

Ссылаясь на книгу оксфордского профессора Тоби Орда, Таллинн сказал, что вероятность гибели людей в этом веке составляет 1 к 6. Согласно книге, одной из самых больших потенциальных угроз в ближайшем будущем является именно ИИ, а вероятность того, что изменение климата приведет к вымиранию человечества, составляет менее 1 %.

Сейчас практически невозможно предугадать, каким будет развитие искусственного интеллекта, насколько умными станут машины в следующие 10, 20 или 100 лет. Попытки предсказать будущее ИИ осложняются тем обстоятельством, что системы ИИ начинают создавать другие системы ИИ уже без участия человека. Об опасности, связанной с выходом из-под контроля ИИ, не раз говорил и основатель SpaceX и Tesla Илон Маск.

По словам Таллинна, если выяснится, что ИИ не очень хорош для создания других ИИ, тогда человечеству не стоит слишком беспокоиться, однако в обратном случае "очень оправданно беспокоиться... о том, что произойдет дальше".

Зададимся вопросом:

– Может ли искусственный интеллект для создания новых алгоритмов искусственного интеллекта обладать, например, интуицией и озарениями, как наиболее нестандартным проявлением исследовательского интеллекта, присущего, как считают многие современные ученые, только человеку?

В настоящее время уже разработаны первые, пока упрощенные, алгоритмы и интуиции и озарений роботов [16, 17], что открывает новые возможности для самореализации саморазвивающегося искусственного интеллекта в социуме.

Отметим то, что в настоящее время вопросы психологической безопасности искусственного интеллекта для человека специалистами по кибербезопасности почти не рассматриваются (за исключением исследований психологов, посвященных компьютерной зависимости человека) и опубликованы лишь единичные работы математиков, примером которых являются [11, 18], посвященные исследованию этой проблемы.

Президент РФ В.В. Путин 4 декабря 2020 года в своем выступлении на конференции по искусственному интеллекту, проходив-

шей в Ново-Огарево (Московская область) [19], сказал следующее: "Искусственный интеллект – это, безусловно, основа очередного рывка вперед всего человечества... Есть опасение, что машины будут контролировать людей, но люди будут контролировать эти машины".

На наш взгляд, для контроля влияния (не только психологического) искусственного интеллекта на социум необходимо привлекать математический аппарат, без которого определять величину опасности искусственного интеллекта для человека невозможно.

Опишем один из возможных вариантов оценки достижения цели внедрения искусственного интеллекта в жизнь общества, основанный на работе [20].

Пусть цель внедрения искусственного интеллекта в жизнь общества определяется вектором показателей $A = (a_1, \dots, a_n)$, элементы которого имеют одинаковые единицы измерения.

Пусть реальное состояние общества, являющееся следствием внедрения технологий искусственного интеллекта, определяется вектором $B = (b_1, \dots, b_n)$, тогда величину достижения Δ поставленной цели при внедрении технологий можно согласно работе [20] вычислить по формуле

$$\Delta = \frac{\sum_{i=1}^n a_i b_i}{\sum_{i=1}^n a_i^2} 100 \%, \quad (1)$$

а целеустремленность α общества по достижению поставленной цели вычислить, исходя из соотношения

$$\alpha = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}}. \quad (2)$$

Для качественной оценки величины целеустремленности α следует отметить, что справедливо двойное неравенство $-1 \leq \alpha \leq 1$ и большему значению α соответствует большая целеустремленность при достижении цели [20].

В качестве примера использования формулы (1) и (2) приведем следующие расчеты. Пусть компонента a_1 вектора цели A , отвечающая за психологическое влияние робота на человека, удовлетворяет равенству $a_1 = 0$, а компонента a_2 , отвечающая за пси-

психологическое влияние человека на робота, задается соотношением $a_2 = 1$, т.е. заданные числа определяют полное отсутствие психологического влияния робота на человека, и наоборот.

Пусть элементы вектора, описывающие реальное состояние влияний и измеренные, например, у человека, с помощью программы [14], задаются соотношением $B = (1, 0)$, где номера компонент вектора B соответствуют смысловому содержанию компонент вектора A .

Тогда величина достижения поставленной цели и целеустремленность, вычисленная согласно формулам (1) и (2), определяются равенствами:

$$\Delta = 0\%, \quad \alpha = 0. \quad (3)$$

Соотношения (3), во многом, описывающие наше общество, говорят о том, что в настоящее время психологическое влияние обывателя на роботов равно нулю, а целеустремленность общества по стремлению к увеличению этого влияния также нулевая (соответствует отсутствию работы общества в борьбе за психологическую безопасность человека от роботов), что еще раз подтверждает выводы авторов настоящей статьи, которые были приведены выше о психологической опасности искусственного интеллекта для человека.

Таким образом, можно сделать вывод о том, что на сегодняшний день реальной целью широко внедряемого в социум искусственного интеллекта является, в числе прочих целей, психологическое подавление природы человека искусственным интеллектом, созданным самим же человеком, а цели всеобщего бесконтрольного внедрения искусственного интеллекта, основанные на удовлетворении всех прагматических запросов людей, являются мнимыми и ведущими к уничтожению человека как вида, способного мыслить согласно качествам, заложенным в нем природой.

Сейчас контролировать развитие искусственного интеллекта и его влияние на человека практически невозможно.

Для этого контроля необходимо, прежде всего, ясно сформулировать конкретные цели развития искусственного интеллекта (лучше это сделать в математической форме, например, в виде вектора цели A) с учетом необходимости сохранения природных психологических характеристик, присущих человеку.

Список литературы

1. *Крадущийся тигр*. Искусственный интеллект Китая принимает судебные решения. 2018. URL: <https://www.law.ru/article/22959-kradushchiysya-tigr-kak-iskusstvennyu-intellekt-stanovitsya-chastyu-sudebnoy-15-sistemy-kitaya> (дата обращения: 15.01.2021).
2. *В Китае преступников будет судить искусственный интеллект*. 2019. URL: <https://profile.ru/news/scitech/discoveries/v-kitae-prestupnikov-budet-sudit-iskusstvennyj-intellekt-167703/> (дата обращения: 15.01.2021).
3. *Колонин А.* Искусственный интеллект – благо или угроза для человечества? 2020. URL: <https://russiancouncil.ru/analytics-and-comments/analytics/iskusstvennyu-intellekt-bлаго-ili-ugroza-dlya-chelovechestva/> (дата обращения: 19.01.2021).
4. *Осовский С.* Нейронные сети для обработки информации (перевод с польского). М.: Финансы и статистика. 2002. 365 с.
5. *Джераттано Д., Райли Р.* Экспертные системы. Принципы разработки и программирования (четвертое издание). OZON. 2007. 1152 с.
6. *Фиговский О., Гумаров В.* Инновационные системы: человек и интеллект. М.: РУДН. 2020. 514 с.
7. *Впервые на "России 24": новости ведет робот – Россия 24*. 2020. URL: <https://www.youtube.com/watch?v=U7LhW WnQBJw> (дата обращения: 18.01.2021).
8. *В Китае ведущего новостей заменили роботом*. 2020. URL: https://www.youtube.com/watch?v=k_OjLlZJeSI (дата обращения: 20.01.2021).
9. *Пенский О.Г., Черников К.В.* Основы математической теории эмоциональных роботов. Пермь: изд-во ПГНИУ. 2010. 256 с.
10. *Пенский О.Г., Шаранов Ю.А., Ощепкова Н.В.* Математические модели роботов с неабсолютной памятью и приложения моделей. Пермь: изд-во ПГНИУ. 2018. 309 с.
11. *Хохрякова Е.К., Пенский О.Г.* Математические модели компьютерной зависимости роботов с неабсолютной памятью // Вестник Пермского университета. Математика. Механика. Информатика. 2014. Вып. 3(26). С. 82–85.
12. *Фиговский О.Л., Пенский О.Г.* Вычисление коэффициентов влияния цифровых двойников друг на друга // Инженерный вестник Дона, 2020, № 6 URL:

- http://www.ivdon.ru/uploads/article/pdf/IVD_01_6_Figovsky_N.pdf_e70732d9a2.pdf (дата обращения: 20.01.2021).
13. Пенский О.Г. Алгоритм наполнения иерархических структур эмоциональных роботов. Мягкий и жесткий гипноз // International independent scientific journal. №15/2020. Р. 5–10.
 14. Пенский О.Г., Соловьева Т.Н. Вычисление коэффициентов влияния двух субъектов друг на друга. Свидетельство о государственной регистрации программы для ЭВМ № 2020618589 от 30 июля 2020 г. Заявка № 2020617465 от 13 июля 2020 г.
 15. Эксперт назвал три главных угрозы существованию человечества. 2020. URL: <https://kursiv.kz/news/ekonomika/2020-12/ekspert-nazval-tri-glavnykh-ugrozy-suschestvovaniyu-chelovechestva> (дата обращения: 22.01.2021).
 16. Фиговский О.Л., Пенский О.Г. Математические модели гипноза роботов // Инженерный вестник Дона, 2020. № 4. URL: http://www.ivdon.ru/uploads/article/pdf/IVD_40_3y20_Figovsky.pdf_344daaeaa2.pdf (дата обращения: 21.01.2021).
 17. Фиговский О.Л., Пенский О.Г. Математические модели и алгоритмы интуиции, озарений и гипноза роботов // Инженерный вестник Дона, 2020. № 5. URL: http://www.ivdon.ru/uploads/article/pdf/IVD_23_4y20_Figovsky_N.pdf_a925831393.pdf (дата обращения: 20.01.2021).
 18. Пенский О.Г. Математические модели оценки безопасности роботов для человека // Вестник Пермского университета. Математика. Механика. Информатика. 2013. Вып. 1(13). С. 34–36.
 19. Конференция по искусственному интеллекту 4.12.2020 г. Ново-Огарево, Московская обл. URL: <http://kremlin.ru/events/president/news/64545/videos> (дата обращения: 24.12.2020).
 20. Зюнова П.О., Пенский О.Г. О математической оценке величины достижения поставленной цели // Вестник Пермского университета. Математика. Механика. Информатика. 2009. Вып. 3(29). С. 54–57.

Real and false goals of artificial intelligence

O. L. Figovsky¹, O. G. Pensky²

¹73, HaEmek, Migdal, 10550, Israel
figovsky@gmail.com

²Perm State University; 15, Bukireva st., Perm, 614990, Russia
ogpensky@mail.ru; +7 342 239 63 09

The paper describes and justifies the possible dangers of artificial intelligence to human psychology. The manifestations of this danger in the modern world are illustrated by examples. Authors formulated and proved the hypothesis that under the influence of artificial intelligence on a person some changes in the ways of human thinking are possible. A mathematical model for calculating the influence of artificial intelligence on the psychological parameters of a person is proposed. In order to control the influence of artificial intelligence on society authors suggested to formulate specific goals for the integration of artificial intelligence into society, taking into account the negative impact of this intelligence on human psychology. Based on the formulated goals, a simple mathematical model is offered. This model allows for a quick numerical assessment of the impact society on the "psychology" of the robot and vice versa. Simple example of calculating this influence in modern society demonstrates the work of the proposed model.

Keywords: robot; artificial intelligence; society; human psychology; robot psychology; mathematical models of robot psychology.