

---

---

# СОЦИАЛЬНАЯ ФИЛОСОФИЯ И ФИЛОСОФИЯ ИСТОРИИ

Л. Е. ГРИНИН, А. Л. ГРИНИН, И. Л. ГРИНИН

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: РАЗВИТИЕ И ТРЕВОГИ. ВЗГЛЯД В БУДУЩЕЕ

### Статья вторая. Искусственный интеллект: terra incognita или управляемая сила?\*

*Работа посвящена истории развития информационно-коммуникационных технологий (ИКТ) и искусственного интеллекта (ИИ), их современным и предполагаемым будущим достижениям в связи с развитием этих технологий и их активным использованием в обществе. Острые проблемы, связанные с ИИ, отчетливо ощущаются уже сейчас, но в будущем они станут намного острее. Показываются тесная связь развития ИИ с когнитивной наукой, мощное внедрение ИКТ и ИИ в различные сферы, в частности сферу охраны здоровья, а также в очень интимные области, связанные с созданием цифровых копий умерших и посмертных контактов с ними. Значительное место в статье уделяется анализу понятия «искусственный интеллект», в том числе определениям общего и генеративного ИИ. Дается анализ последних достижений в области искусственного интеллекта, описываются базовые модели, в частности большие языковые модели (large linguistic models, LLM). Предлагаются*

---

\* Исследование выполнено при поддержке Российского научного фонда (проект № 23-11-00160 «Моделирование и прогнозирование развития стран БРИКС в XXI веке в контексте мировой динамики»).

**Для цитирования:** Гринин Л. Е., Гринин А. Л., Гринин И. Л. Искусственный интеллект: развитие и тревоги. Взгляд в будущее. Статья вторая. Искусственный интеллект: terra incognita или управляемая сила? // Философия и общество. 2023. № 4. С. 5–32. DOI: 10.30884/jfio/2023.04.01.

**For citation:** Grinin L. E., Grinin A. L., Grinin I. L. Artificial Intelligence: Development and Concerns. A Look into the Future. Article two. Artificial Intelligence: Terra Incognita or a Controlled Force? // Filosofiya i obshchestvo = Philosophy and Society. 2023. No. 4. Pp. 5–32. DOI: 10.30884/jfio/2023.04.01 (in Russian).

*Философия и общество, № 4 2023 5–32*

*DOI: 10.30884/jfio/2023.04.01*

прогнозы развития ИИ и опасностей, которые будут подстерегать нас в ближайшие десятилетия. Указывается, какие силы стоят за стремлением создать могущественный искусственный интеллект, все более приближающийся по своим возможностям к так называемому общему / универсальному ИИ, а также обсуждаются желательные действия по ограничению и канализации развития искусственного интеллекта. Подчеркивается, что угрозы и опасности развития ИКТ и ИИ особенно усиливаются в связи с монополизацией их развития со стороны государства, спецслужб, крупнейших корпораций и тех, кого нередко называют глобалистами. В статье даны прогнозы развития компьютеров, ИКТ и ИИ в ближайшие десятилетия, а также показаны изменения в обществе, которые будут с этим связаны.

Работа состоит из двух статей. В первой из них, опубликованной в предыдущем номере журнала, была описана краткая история и дана характеристика текущей ситуации в области ИКТ и ИИ, анализировались понятия искусственного интеллекта, в том числе генеративного ИИ, изменений в понимании ИИ в связи с появлением так называемых больших языковых моделей и на их базе – новых типов ИИ (ChatGPT и аналогичные модели). В статье обсуждались серьезные проблемы и опасности, связанные с быстрым и неконтролируемым развитием искусственного интеллекта.

В настоящей, второй статье описываются и комментируются современные оценки прорыва в области ИИ, анализируются различные прогнозы, авторы дают собственные оценки и прогнозы будущего развития. Особое внимание уделено проблемам и опасностям, связанным с быстрым и неконтролируемым развитием ИИ, с тем, что достижения в этой области становятся мощнейшим средством контроля над населением, навязывания идеологии, приоритетов и стиля жизни, способом влиять на результаты выборов, орудием подрыва безопасности и геополитической борьбы.

**Ключевые слова:** информационно-коммуникационные технологии, ИКТ, искусственный интеллект, ИИ, большие языковые модели, LLM, когнитивная наука, самоуправляемые системы, кибернетическая революция, инфорг, технологический прогресс.

*The article is devoted to the history of the development of ICT and AI, their current and expected future achievements, and the problems (which have already arisen but will become even more acute in the future) associated with the development of these technologies and their widespread application in society. It shows the close connection between the development of AI and cognitive science, the penetration of ICT and AI into various spheres, particularly health care, and the very intimate areas related to the creation of digital copies of the deceased and posthumous contact with them. A significant part of the article is*

*devoted to the analysis of the concept of “artificial intelligence”, including the definition of generative AI. The authors analyse recent achievements in the field of Artificial Intelligence. There are given descriptions of the basic models, in particular the Large Linguistic Models (LLM), and forecasts of the development of AI and the dangers that await us in the coming decades. The authors identify the forces behind the aspiration to create AI, which is increasingly approaching the capabilities of the so-called general/universal AI, and also suggest desirable measures to limit and channel the development of artificial intelligence. It is emphasized that the threats and dangers of the development of ICT and AI are particularly aggravated by the monopolization of their development by the state, intelligence services, major corporations and those often referred to as globalists. The article provides forecasts of the development of computers, ICT and AI in the coming decades, and also shows the changes in society that will be associated with them.*

*The study consists of two articles. The first, published in the previous issue of the journal, provided a brief historical overview and characterized the current situation in the field of ICT and AI. It also analyzed the concepts of artificial intelligence, including generative AI, changes in the understanding of AI in connection with the emergence of the so-called large language models and related new types of AI programs (ChatGPT and similar models). The article discussed the serious problems and dangers associated with the rapid and uncontrolled development of artificial intelligence.*

*This second article describes and comments on current assessments of breakthroughs in the field of AI, analyzes various predictions, and provides the authors’ own assessments and predictions of future developments. Particular attention is paid to the problems and dangers associated with the rapid and uncontrolled development of AI, with the fact that advances in this field are becoming a powerful means of control over the population, imposing ideology, priorities and lifestyles, influencing the results of elections, and a tool to undermine security and geopolitical struggles.*

**Keywords:** *information and communication technologies, ICT, artificial intelligence, AI, large language models, LLM, cognitive science, self-regulating systems, the Cybernetic Revolution, inforg, technological progress.*

### **1. Симбиоз когнитивных дисциплин и ИИ**

Развитие ИКТ и ИИ во многих случаях ярко демонстрирует многообразную связь, а также взаимозависимость, переходящую в теснейший симбиоз передовых технологий МАНБРИК [примеры см.: Гринин Л. Е., Гринин А. Л. 2015а; 2015б; Grinin L., Grinin A. 2016; Grinin *et al.* 2017а; 2017б; 2020; 2021]<sup>1</sup>. В частности, налицо

---

<sup>1</sup> МАНБРИК – это аббревиатура первых букв семи направлений, которые, как мы считаем, станут играть ведущую роль в будущем: медицина, аддитивные тех-

комбинация нейротехнологий, с одной стороны, ИКТ и ИИ – с другой, когда разработчики искусственного интеллекта стремятся использовать достижения когнитивной науки для его развития. Дело в том, что по крайней мере в последние три десятилетия именно достижения когнитивных наук стали одним из важных драйверов ИИ. Мы имеем в виду быстро развивающуюся технологию нейронных сетей (точнее, сейчас они называются нейроморфными сетями, а область их изучения – нейроморфные вычисления)<sup>2</sup>. Для развития ИИ нужно было понять, как работает человеческий мозг [см., например: Hawkins, Blakeslee 2004]. Недостаток знаний о строении и работе мозга, механизмах памяти, принятия решений, предвидении и других интеллектуальных функциях мозга на глубинных уровнях стал препятствием для продвижения по пути машинного обучения и других возможностей. В итоге начал формироваться симбиоз когнитивных наук и технологий, с одной стороны, и программирования – с другой, на пути изучения работы мозга с тем, чтобы использовать это для технологий машинного обучения, а также расширения возможностей влиять на человеческое сознание с помощью ИИ. Такой подход сулил большие перспективы, в результате чего стал активно поощряться и возглавляться правительственными структурами. Считается, что ключ к дальнейшим успехам лежит в сборе, накоплении, обработке, анализе и использовании данных о мозге. В США развернуты масштабные программы сбора и анализа данных о мозге, но то же самое имеет место в европейских государствах, Китае и других странах. Очень показательным в этом плане стал проект Apollo Project of the Brain 2016 г., на который американское правительство выделило 100 млн долларов. Цель проекта – найти алгоритмы, которые позволили бы компьютеру думать подобно людям. Агентство передовых исследований в сфере разведки (The Intelligence Advanced Research Projects Activity/IARPA), созданное как аналог знаменитого Агентства по перспективным оборонным научно-исследовательским разработкам

---

нологии, нанотехнологии, биотехнологии, робототехника, ИКТ и ИИ, когнитивные технологии.

<sup>2</sup> При моделировании нейронных сетей разработчики стремятся сделать их максимально похожими с точки зрения обработки информации на биологические нейронные сети. Информация кодируется в таких сетях в виде интервалов между импульсами, генерируемыми нейроном в ответ на локально интегрированное в пространстве и времени возбуждение от импульсных сигналов, поступающих на его входы. Технология ориентируется на полностью аппаратную реализацию [Бендерская, Толстов 2013; Гаврилов, Канглер 2015].

Министерства обороны США (DARPA), также выделило 100 млн долларов на аналогичный грандиозный проект «Искусственный интеллект на основе нейронных сетей» (Machine Intelligence from Cortical Networks / MICrONS) для проведения инженерного анализа образца мозга величиной в один кубический миллиметр, изучения механизмов, с помощью которых мозг выполняет вычисления, и на основе полученных данных – повышения производительности алгоритмов машинного обучения и искусственного интеллекта [Cerelewicz 2016]. Чтобы подчеркнуть всю необычность и объемность этой задачи, приведем данные от Allen Institute [An Automated... 2020], показывающие, что его роль заключалась в осуществлении того, что еще никогда не было сделано: разделении участка коры головного мозга объемом в один кубический миллиметр на ~25 тыс. ультратонких срезов, сделать ~125 млн фотографий этих срезов и собрать их в трехмерный объем, содержащий ~100 тыс. клеток, 2,5 мили проводов и 1 млрд синаптических связей.

Словом, развитие ИИ во многих его направлениях основано на попытках имитировать различные биологические механизмы<sup>3</sup>. Конечно, достичь возможностей мозга искусственному интеллекту очень непросто. Наиболее совершенные нейронные сети имеют сегодня десятки слоев. В человеческом мозге таких слоев сотни тысяч и миллионы. Это, естественно, ограничивает многие возможности, включая и т. н. *deep learning* [Schmidhuber 2014; Гаврилов, Канглер 2015]. Тем не менее даже на первых шагах успехи ИИ являются очень впечатляющими и одновременно тревожными, так как в первую очередь ими собираются воспользоваться структуры, намерения и технологии которых секретны и неподотчетны. «Это существенные инвестиции, поскольку мы считаем, что это серьезная задача, и [она окажет] преобразующее воздействие на разведывательное сообщество», – говорилось в комментариях к вышеописанному проекту [Cerelewicz 2016].

---

<sup>3</sup> Особенностью нейроморфных сетей является возможность моделирования фрагментов биологической нервной системы и интеллектуальных свойств мозга; эволюционные вычисления моделируют естественную эволюцию и формализуют основные законы генетики; роевой интеллект моделирует социальное поведение организмов, которые живут колонией (роем, стаей и т. п.); искусственные иммунные системы моделируют основные принципы биологических иммунных систем; нечеткие системы основаны на исследованиях взаимодействия организмов с окружающей средой [Скобцов 2008].

## 2. Шаг к общему ИИ и возможные опасности<sup>4</sup>

Современный продвинутой ИИ часто называют генеративным. Подробно мы говорили об этом в первой статье. Однако стоит напомнить, что генеративный ИИ – это такой тип искусственного интеллекта, который может генерировать новое содержание, формулировать идеи, вести диалоги, создавать произведения, истории, картинку, видео и музыку, а также редактировать изображения, видео и пр. Он основан на современных базовых моделях обучения ИИ, которые называют большими языковыми моделями (LLM). Они специально ориентированы на выполнение языковых задач, включая создание (генерацию) текстов, информации, блогов, диалогов, извлечение информации. Мощные генеративные ИИ «тренирует» на сверхдорогих суперкомпьютерах команда высококлассных программистов и других специалистов, включая психологов. Для их обучения используют невероятные объемы информации из самых разных областей.

**Тема искусственного интеллекта становится предметом широкого общественного обсуждения.** Опасности со стороны ИИ и злоупотребления им давно были видны не только специалистам<sup>5</sup>. Его влияние (часто негативное и сильно раздражающее) мы ощущаем практически каждый день (см. также примеры в первой статье). Немало писалось и об этике, связанной с развитием и использованием ИИ [см., например: Разин 2019]. Но работа в области такой этики требует аккумуляции и концентрации сил всего общества, а также вполне четких императивов и правил. Пример можно

---

<sup>4</sup> Напомним, что общий (сильный, универсальный) ИИ – это гипотетический ИИ в философии искусственного интеллекта, способный выполнить любую задачу, с которой справляются люди (подробно об этом мы говорили в первой статье). Пока это гипотеза и до сих пор неясно, возможно ли такое в принципе, но продвижение в данном направлении происходит гораздо быстрее, чем предполагалось ранее, и эта скорость в дальнейшем может еще очень сильно возрасти. Отсюда опасность неконтролируемого использования все более мощного и изощренного ИИ.

<sup>5</sup> Указывалось, что быстрое развитие технологии искусственного интеллекта представляет собой главный вызов для человечества [Singularity... 2012]. С тех пор как об этом заговорили, коммуникационные технологии, технологии анализа данных и наблюдения продвинулись очень значительно, даже радикально. В результате проблема стала еще более актуальной. Анализ различных аспектов этой проблемы в настоящем и будущем уже посвящен целый ряд работ [см., например: Westin 1966; Ashman *et al.* 2014; Cecere *et al.* 2015; Moustaka *et al.* 2019; Schwartz 1999; Solove 2008; Brammer *et al.* 2020; Alharbi 2020; Гринин Л. Е., Гринин А. Л. 2023].

брать с широко известных постулатов, сформулированных в научно-фантастических произведениях Айзека Азимова «Я, робот» и других книгах из серии о роботах. Тем не менее проблема влияния развития ИИ на общество, на человечество, и тем более на будущее человечества как расы; того, что под угрозой оказывается его эволюция, никогда не стояла в центре внимания. В 2023 г. все изменилось, правда, во многом именно потому, что силы, которые стоят за развитием ИИ, были заинтересованы в нагнетании хайпа и сенсационности вокруг этой темы по причинам, связанным с желанием дать импульс росту акций технологических компаний (подробнее об этом мы скажем ниже, однако ситуация весьма похожа на ту, которая сложилась в отношении пандемии и вакцин от COVID-19). Но, так или иначе, данная тема стала одной из ведущих в общественной повестке – и это, как мы полагаем, очень хорошо.

**ChatGPT: достижения.** Одним из последних примеров сложнейшего ИИ является появившаяся на свет в 2020 г. текстовая нейросеть GPT-3 (она относится и к ИИ больших языковых моделей). Став преемницей GPT-2, после долгих лет разработки она являла собой нечто поистине устрашающее – около 200 млрд параметров внутренней разработки, стоимость обучения в 5 млн долларов и 570 Гб текста для обучения (ее предшественница GPT-2 обучалась «только» на 40 Гб данных [Каспарьянц 2022]). Компания-разработчик Open AI выдала права на пользование ею лишь нескольким крупным платформам, однако спустя три года представила ChatGPT – чат-бот, основанный на возможностях самой нейросети. Его возможности поражают воображение. Очевидно, в относительно ближайшем будущем (в пределах 5–15 лет) ChatGPT начнет активно заменять человека в плане интеллектуального труда, в частности написания текста во многих областях (журнализм, реклама; даже медицина и научные тексты и т. д.; см.: Biswas 2023; Lund, Ting Wang 2023]. Об этом мы также поговорим ниже<sup>6</sup>. Например, появилась информация, что Amazon уже стал рынком для томов, созданных искусственным интеллектом, которые выдаются за написанные людьми, причем книги о путешествиях входят

---

<sup>6</sup> Ажиотажа вокруг этого чата очень много, от апокалиптических прогнозов до нигилизма (подробнее см. ниже). Например, некоторые современные статистические данные показывают, что на данный момент ChatGPT «глупеет» в связи с большим количеством запросов и процентная верность его ответов снижается. Окажет ли это влияние на будущее развитие или данный недостаток возможно исправить – пока не совсем ясно.

в число популярных категорий поддельных работ. Полагаем все-таки, что через некоторое время таких «сочинителей» обяжут писать, что произведение создано ИИ.

Open AI, создавшая ChatGPT, также продемонстрировала нейросеть DALL-E 2, которая способна превращать текстовые описания в фотореалистичные и художественные изображения. Она иллюстрирует стихи Бродского, дорисовывает картины классиков. Она также способна создавать изображения даже из бессмысленных (но грамматически правильных) предложений, в частности, для известного примера подобного предложения Хомского: «Бесцветные зеленые идеи яростно спят». При этом психологические эксперименты психологов Гарварда доказали полное отсутствие у DALL-E 2 даже намека на понимание результатов своей деятельности, многими людьми ошибочно воспринимаемых как творчество ИИ. Тем не менее такие возможности в любом случае повлекут за собой очень серьезные изменения для людей творческих профессий и для общества в целом.

Если такого рода чат-боты станут очень удобными секретарями-референтами и интеллектуальными помощниками, «советчиками» и «советниками», наставниками в тех или иных операциях, разрешении проблем и т. п., это может значительно облегчить подготовительную работу во многих областях интеллектуальной деятельности. Но есть и серьезная опасность, что люди перестанут перепроверять данные и будут полагаться на ответы ИИ, что понизит уровень глубины, а также достоверности, не говоря уже об обобщении личного опыта и неизбежных серьезных ошибках, винить за которые придется только себя. В целом это процесс, который идет уже давно (в качестве примера можно привести «Википедию»), но теперь он ускорится. Люди неизбежно станут лениться, и их стиль начнет следовать за стилем чата, в который можно помещать то, что посчитают нужным спонсоры и хозяева. То есть ИИ начнет диктовать стиль мышления, изложения и изображения. Рассмотрим возможности современного искусственного интеллекта на примере<sup>7</sup>.

**ИИ как помощник и как угроза.** Очевидно, что во многих отношениях использование больших языковых моделей позволит сделать работу различных специалистов гораздо более продуктивной, даст возможность проводить такие изыскания, расчеты и ис-

---

<sup>7</sup> При создании этой статьи нам очень помогла информация телеграм-канала «Малоизвестное интересное» [Малоизвестное...].



следования, которые в сегодняшних условиях просто немыслимы. Хороший пример такого прорыва в производительности ученого представил Патрик Мино, который изучает способы преодоления схизмогенеза<sup>8</sup> в рамках исследований в области NeuroAI. Он стремится в том числе понять, какой из путей исследований более перспективен:

- использование результатов нейробиологии для совершенствования ИИ (нейро → ИИ);
- или наоборот (ИИ → нейро).

В рамках решения этой задачи было необходимо выявить исследования, касающиеся взаимосвязей и взаимовлияния нейробиологии и ИИ. Для этого были задействованы большие языковые модели. В результате ИИ: а) сначала проанализировал 40 тыс. научных статей по нейробиологии и ИИ за последние 40 лет; б) затем выявил из 40 тыс. 1,5 тыс. статей, касающихся взаимосвязей и взаимовлияния нейробиологии и ИИ; в) на третьем этапе по результатам анализа расположил выявленные 1,5 тыс. статей в придуманном Мино ландшафте. Естественно, что провести подобный анализ силами не только одного человека, но даже отдела, а возможно, и института, практически немыслимо.

Но в этом отношении большие языковые модели – просто еще один, хотя и заметный, шаг вперед, начиная с первых ЭВМ, которые позволили за считанные дни делать расчеты, на которые ранее могли уйти долгие годы. Или возьмем, например, Excel. Программа позволяет за секунды производить вычисления и создавать математические модели, на которые раньше ушли бы годы. При этом такие модели может делать и не математик, чего раньше просто невозможно было себе представить.

Налицо процесс, о котором мы неоднократно писали – переход от узкой специализации, которая развивалась в течение многих веков, к универсализации компетенций и навыков за счет самоуправляемых систем и искусственного интеллекта [см., например: Гринин Л. Е., Гринин А. Л. 2015б].

При этом очень вероятно, что новый, еще более продвинутый ИИ очень существенно (а может быть, и радикально именно в этом отношении) повлияет на программистов за счет того, что с помощью подобных чатов программирование значительно облегчится, а соответственно, может снизиться как потребность во все большем

---

<sup>8</sup> Схизмогенез – это изменение индивидуального поведения, происходящее в результате накопления опыта взаимодействия между индивидами.

количестве программистов, так и уровень доходов большинства из них. Разумеется, это процесс, который растянется на десятилетия, однако мы боимся повторения старой истории, когда создание пожирает своего создателя. Главное, конечно, чтобы ИИ не превратился в чудовище Франкенштейна.

Так или иначе, можно согласиться, что через 10–20 лет личные LLM-помощники многократно превзойдут умных помощников Алису (Яндекс), Siri (Apple) или Alexa (Амазон) и станут повсеместными [см.: Chat... 2023]. Эти помощники обретут почти сказочные возможности, по крайней мере в области интеллектуальной помощи и информации. Получится, что многие или даже каждый из нас смогут иметь у себя «Двоих из ларца» или «чудище» из сказки «Аленький цветочек» (см. о возможностях помощников выше). И это неудивительно, ведь многие сказки человек уже воплотил в реальность. Однако крайне важно, чтобы помощники не превратились в орудие, с помощью которого нас будут постоянно и все более глубоко загонять в виртуальный концлагерь.

**Опасности.** Таким образом, мы выходим на новый уровень внедрения в нашу интеллектуальную жизнь новых возможностей ИИ. Это будет требовать серьезной трансформации образования, а также искусства, науки (которые, конечно, частично будут деградировать под таким напором), а также других сфер. В итоге чудовищное количество искусственно созданной интеллектуальной продукции будет превалировать над ее качеством.

Еще бо́льшая опасность в том, что такие чаты немедленно начнут использовать как опосредованную мощнейшую идеологическую и пропагандистскую машину (чем большим количеством соответствующих текстов с нужным уклоном наполнить память, тем больше будет пропагандистский эффект), также они могут служить как мощный механизм цензурирования. Вероятно, со временем начнется и борьба идеологических чатов друг с другом, как сейчас идет борьба сайтов. Особенно серьезная опасность развития ИИ (как, впрочем, и других технологий) в том, что мы осознаем эти угрозы только тогда, когда изменить что-то оказывается очень непросто. Как заметил Билл Гейтс, «какие бы ограничения он (ИИ) ни имел сегодня, они исчезнут, прежде чем мы об этом узнаем» [Gates 2023]. Поэтому регламентации должны быть сделаны заранее, заблаговременно.

Говоря об опасениях в отношении новейшего ИИ (типа ChatGPT или DALL-E 2), следует заметить, что разобраться в оценках этого

явления сейчас очень сложно, поскольку никакого взвешенного и систематизированного взгляда, а тем более внятной теории здесь не просматривается. Зато очень много истерики, восхищения, чепухи и хайпа. Немедленно заговорили о «революции ChatGPT», хотя революция это или просто новый рывок информационных технологий, каких мы пережили уже с десятком, совершенно не ясно. Во многом (если не в основном) этот ажиотаж связан с попытками разогнать рынок, как бывало и ранее с прорывными технологиями, в частности с нанотехнологиями [подробнее см.: Гринин Л. Е., Гринин А. Л. 2015a]. Мы стали свидетелями мощной информационной волны, общий смысл которой: новый ИИ не даст экономике спуститься в рецессию<sup>9</sup>, в результате чего без фундаментальных причин акции ведущих технологических компаний, а в результате и в целом рынки акций в 2023 г. поднялись весьма значительно<sup>10</sup>. Таким образом, застрельщикам этой медиакампании удалось поднять интерес к акциям цифровых компаний.

Считается, что развитие использования ИИ ведет к адаптации технологий в классических отраслях экономики по всей цепочке создания ценности и преобразует их, приводя к алгоритмизированию практически всего функционала, от логистики до управления компанией, от выбора ценовой политики до анализа рынков. В целом это так. Но в ситуации все более сильной финансиализации западной экономики компании прежде всего стремятся заработать

---

<sup>9</sup> Например, рынок только генеративного ИИ (это ChatGPT и другие чат-боты) может вырасти в 30 раз к 2032 г. и достичь 1,3 трлн долларов [см., например: Rudnitsky 2023].

<sup>10</sup> В итоге компании, которые ссылались на ИИ в отчетах о прибылях и убытках за второй квартал, демонстрировали, по расчетам FACTSET, лучшую динамику цен на акции, чем те, что не ссылались.

Это раздувание за счет ИИ, похожее на аферу, продолжается уже более полугода. Приведем только один пример: 12 сентября 2023 г. наблюдался взлет акций компании Tesla, которую аналитики Morgan Stanley в своем обзоре накрепко связали с искусственным интеллектом. Компания намерена потратить более \$1 млрд на суперкомпьютер Dojo для обучения ИИ на беспилотных автомобилях. Аналитики сочли, что это откроет перед ней новые масштабные рынки и повысит капитализацию на 600 млрд долларов.

До последнего пока далеко, но 80 млрд долларов Tesla прибавила в тот же день. Взлет ее акций более чем на 10 % был самым мощным с января этого года. В этом смысле ажиотаж вокруг ИИ напоминает другую известную манию – «пузырь доткомов» в 1990-х гг. Тогда любая компания, в названии которой появлялось что-нибудь вроде «\*.com», могла рассчитывать на положительную переоценку в несколько раз, если не на порядок [Баранов 2023].

на хайпе, медийных кампаниях и надувании пузырей на рынке акций. Появилось такое сообщение. По данным Goldman Sachs, генеративный ИИ может обеспечить увеличение мирового ВВП на 7 % (или почти на 7 трлн долларов) и поднять рост производительности на 1,5 п. п. в течение 10 лет [Что такое... 2023]. Разумеется, это слишком большие цифры, которые являются не столько прогнозом, сколько возможностью захватить воображение инвесторов и побудить их вложить средства в акции технологических компаний (см. выше о росте акций Tesla). Оптимистичным было бы считать, что за 10 лет генеративный ИИ сможет добавить в мировой ВВП 1 %, но и это будут очень большие цифры.

Высказываются мнения, что область применения генИИ повсеместна, как с электричеством [Fraleu 2023]. Безусловно, спектр применения генИИ будет расширяться, но сравнение с электричеством выглядит, конечно, серьезным преувеличением.

### **3. Страхи и опасности: реальные и мнимые**

#### **3.1. Наиболее часто обсуждаемые страхи**

**Многочисленные страхи относительно ИИ** можно разделить на несколько направлений:

1. ИИ подчинит себе людей.
2. ИИ лишит нас рабочих мест.
3. Скорость изменений ИИ начнет расти на порядок.
4. Примерно через 10 лет ИИ заменит преподавателей.

Теперь рассмотрим их подробнее.

**ИИ подчинит себе людей.** Вот пример таких взглядов – «ИИ превращает человечество в очень маленькое явление по сравнению с чем-то гораздо более разумным, чем мы, и ИИ станет для нас непостижимым, как мы для тараканов» (цитата из интервью американского информатика Дугласа Хофштадтера [Gödel... 2023]). Стоит привести и другие его высказывания:

«Для меня это (прогресс ИИ) довольно страшно, потому что это говорит о том, что все, во что я раньше верил, отменяется. Я думал, что пройдут сотни лет, прежде чем появится что-то хотя бы отдаленно похожее на человеческий разум».

«Я никогда не представлял, что компьютеры будут соперничать или даже превосходить человеческий интеллект. Это была настолько далекая цель, что я не беспокоился об этом. А потом это начало происходить все быстрее, когда недостижимые цели и вещи, кото-

рые компьютеры не должны были быть в состоянии сделать», начали осуществляться и происходить.

«Прогресс ИИ ужасает. Я ненавижу его. Я думаю об этом практически все время, каждый божий день».

«Такое ощущение, что вся человеческая раса вот-вот будет превзойдена и оставлена в пыли».

«Ускоряющийся прогресс оказался настолько неожиданным, что застал врасплох не только меня, но и многих, многих людей – возникает некий ужас перед надвигающимся цунами, которое застанет врасплох все человечество».

Все это выглядит как технологическое кликушество. Нет, мы не отрицаем возможности попыток поработить человечество со стороны группы могущественных людей, называемых нами глобалистами, а также национальных правительств и крупнейших корпораций с помощью целенаправленного и ускоренного развития ИИ и мощности суперкомпьютеров. Этот процесс идет уже десятилетия и вполне может ускориться. Однако технологии сами по себе никого не поработают и не закрепощают. В худшем случае они заставляют общество и людей несколько перестраиваться, менять образ жизни, квалификацию и пр., но только тогда, когда они реально дают мощный подъем производительности труда, экономии ресурсов (либо вовлечения в оборот новых ресурсов) и роста богатства. Главная мысль: любая технология подконтрольна тем, кто ею управляет. Поэтому опасность исходит от тех групп, которые могут обрести чрезмерную власть и неконтролируемые возможности с помощью ИИ. Эти группы нужно разоблачать и ограничивать. А кликушество о том, что ИИ сам по себе поработит человечество и «обратит его в пыль», только затемняет истинные опасности процесса. Кроме того, опыт показывает, что опасные технологии в результате противодействия им (которое крайне необходимо) как-то уравниваются<sup>11</sup>. Поэтому важно определить как глубинные причины (просим извинить за каламбур) искусственного подстегивания развития ИИ, так и те интересы и силы, которые за этим стоят.

---

<sup>11</sup> Это произошло даже с такой страшной угрозой, как ядерное оружие. Взаимная опасность противоборствующих сторон снизила риск его применения, хотя, к сожалению, сегодня вновь стали очень много и безответственно говорить о ядерной войне.

**ИИ лишит нас рабочих мест.** Например, согласно аналитическим отчетам компаний Nelson, McKinsey и Pricewaterhouse-Coopers, уже к 2030 г. ИИ заменит от 30 до 50 % существующих сегодня профессий. В таком виде – это сенсационные (в самом негативном смысле этого слова) «страшилки». Однако за ними стоят большие интересы тех политических и финансовых национальных и глобальных сил, которые стремятся использовать новый ИИ, чтобы еще более подчинить себе интеллектуальный процесс миллиардов людей.

Тем не менее и игнорировать опасность подрыва благополучия тех или иных профессиональных групп, в том числе высокоинтеллектуальных, а главное – творческих, было бы неправильно. Уже возникают конфликты, среди которых наиболее известны протесты и забастовки сценаристов Голливуда. Писатели требуют запретить ИИ нарушать их авторские права<sup>12</sup>. В Японии уже было указано, что генеративный ИИ не вправе нарушать копирайт. Иными словами, может появиться запрет на использование чужих произведений для обучения ИИ (что в целом кажется правильным).

Поговорим об угрозах другим профессиям, равно как об опасности, описанной выше, и о проблемах, связанных с необходимостью менять квалификацию, немного подробнее. Разумеется, никаких сокращений профессий до 2030 г. на 30 и тем более на 50 % не будет и быть не может. В последние 15 лет мы часто слышим, что роботы вскоре заменят людей повсюду. Но пока дело обстоит наоборот: на рынке труда острый дефицит работников и предложения рабочей силы. И на первом этапе ИИ, скорее, будет выступать в качестве помощника, а не конкурента – примерно как навигатор облегчил работу таксистов, но не вытеснил их. Тем не менее представителям ряда профессий в перспективе нужно думать о том, что их навыки станут менее востребованными, а со временем им придется потесниться, поскольку ряд их функций возьмет на себя ИИ.

---

<sup>12</sup> Группа писателей из США подала иск к компании OpenAI в федеральный суд Манхэттена. В своем иске писатели обвиняют компанию в том, что она использовала их тексты, защищенные авторскими правами, для обучения программного комплекса искусственного интеллекта ChatGPT [Ott 2023]. Актеры также протестуют против использования их лиц для тренировки нейросети, что способно привести к частичной их замене и сэкономит большие суммы производителям фильмов. Протестуют актеры дубляжа и участники Союза дикторов России, стремящиеся внести в закон ряд поправок, которые помогут защитить их от нелегального синтеза голосов.

Это особенно важно для тех, кто выбирает будущую профессию для себя или своих детей. Не так давно казалось, что знание иностранного языка всегда прокормит человека, но теперь профессия переводчика оказалась под угрозой. ИИ в будущем также способен заменить корректоров, банковских служащих (их число стремительно сокращается уже довольно длительное время), туроператоров, диспетчеров, дикторов, дизайнеров и художников-иллюстраторов, копирайтеров и журналистов, даже писателей и сценаристов (если не удастся запретить использование текстов с копирайтом, см. выше), дикторов, в какой-то мере юристов, учителей, врачей, бухгалтеров, аналитиков и даже, как указывалось выше, программистов (но в этом отношении перспектива более отдаленная). Уже давно созданы магазины без продавцов. Роботы с развитым ИИ в принципе угрожают таксистам и водителям (но это дело тоже не слишком близкого будущего, в пределах 30 лет [см.: Гринин Л. Е., Гринин А. Л. 2015б; 2023]), а также курьерам, уборщикам, нянечкам и гостиничным служащим, поварам, официантам и др.

**Вытеснение профессий: важно, чтобы разрушение было соизмерительным.** Нужно понимать, что прежде всего это сокращение среди творческих профессий и программистов коснется специалистов с достаточно низкой квалификацией. Их ИИ может заменить относительно легко, однако в том, что касается высококлассных специалистов, это будет очень непросто. Так, сегодня сильно поредели ряды фотографов, стало мало фотоателье. Однако хорошие фотографы по-прежнему пользуются спросом. Следовательно, нужно думать о повышении своей квалификации, об особой специализации и нишах, которые ИИ окажется трудно занять (не будет смысла развивать его в этих узких направлениях, см. об этом также в разделе 3.2). Таким образом, упрощение овладения какими-то компетенциями позволит внедриться в ряд областей большому количеству непрофессионалов, могущих стать профессионалами с помощью ИИ.

Кроме того, надо ожидать, что появится много новых профессий и специализаций. Технологический гигант Dell прогнозирует, что 85 % профессий, которые в ближайшие 15 лет будут актуальными, еще даже не существуют. Разумеется, такая цифра приведена для того, чтобы ошарашить публику, но доля истины здесь есть. Некоторые профессии пока еще даже не проглядываются. Здесь, как и при любом технологическом процессе, действует принцип со-

зидательного разрушения, сформулированный экономистом Й. Шумпетером. Но крайне важно, чтобы разрушительная фаза не была слишком тяжелой, жесткой, мучительной и стремительной. Для этого необходимо думать и регулировать сегодня.

Корпорация IBM в своем обзоре «Дополненная работа в автоматизированном мире, управляемом ИИ» [IBM 2023] высказала несколько важных идей (наряду, правда, с шокирующими и, думается, неверными в плане сроков прогнозами).

По оценкам IBM, наступает новая эра в разделении труда между людьми и машинами. Это означает необходимость переквалификации, подразумевает умение встроиться в новое разделение труда между людьми и машинами и освоение навыков «дополненной работы», когда партнерство человека и машины кратно повышает производительность труда и обеспечивает экспоненциальный рост отдачи от бизнеса. Это интересный подход, однако выводы очень настораживают (они явно направлены на повышение курса акций цифровых компаний, см. выше). В *ближайшие три года* 40 % работающих должны будут переквалифицироваться в связи с внедрением ИИ. Это означает, что переквалификацию должны будут пройти 1,4 млрд из 3,4 млрд человек мировой рабочей силы. Примерный процент рабочих мест, которые перейдут к «дополненной работе», составит три четверти в сфере маркетинга (73 %) и обслуживания клиентов (77 %) и более 90 % в сфере закупок (97 %), рисков и соблюдения нормативных требований (93 %) и финансов (93 %).

Повторим, выводы IBM с упомрачительными цифрами переквалификации или дополненной работы, то есть работы, намного более тесно связанной с возможностями ИИ, сделаны именно для ошеломления корпоративных клиентов, чтобы они сразу же пожелали инвестировать свои деньги в искусственный интеллект, то есть в акции компаний, развивающих ИИ. Это цинично и вполне в духе таких корпораций. Разумеется, если такой переход и будет (а он все же, скорее всего, начнется и постепенно станет захватывать все новые сферы), то срок его окажется в разы дольше. Однако в любом случае встает очень важный вопрос, как будет выглядеть эта дополненная работа: станет ли ИИ помощником человека-специалиста или, наоборот, специалист станет придатком ИИ (как в период ранней индустриализации и во времена внедрения конвейерной работы рабочий стал придатком машины, по определению К. Маркса и Ф. Энгельса [1955: 430])?



**Не ИИ заменит людей, а люди, использующие ИИ, заменят тех, которые этого не умеют, не хотят или не могут.** Эта вполне интересная мысль имеет место среди смеси разумных идей и явно провокационных прогнозов, сделанных IBM [2023]. Она важна для понимания того, что ИИ – не самостоятельный субъект, а лишь инструмент (хотя и становящийся все более умным) в руках крупнейших финансистов, цифровиков и разного рода работодателей. Разумеется, ИИ, как раньше машины, а частично и роботы (например, в автомобилестроительной отрасли), во многих профессиях заменяет людей. Но, поскольку это очевидно и стоит на первом плане, то из внимания выпадает второй важный аспект. А именно то, что люди, овладевшие ИИ, оказываются в выигрышном положении по сравнению с теми, кто им не овладел. Следовательно, овладение техническим прогрессом оказывается в приоритете. Для наиболее рьяных поклонников компьютерно-цифровой и интеллектуально-искусственной среды даже появилось особое название *инфорги*, то есть люди, проводящие в цифровой реальности больше времени, чем во сне. Однако идея, что именно инфорги заменят людей, слабо владеющих ИИ, не кажется бесспорной. Для занятия правильной позиции в надвигающемся мире ИИ требуется баланс овладения техникой и наличия собственной аналитики, здравый и критический взгляд на вещи, самостоятельное мышление. Крайне опасно также полностью отдавать творчество на откуп ИИ, иначе это будет уже эрзац творчества невысокого уровня.

**О скорости развития ИИ.** Утверждается, что революция в области искусственного интеллекта достигла переломного момента, и дальше скорость изменений вырастет стремительно, на порядок [IBM 2023]. Мы полагаем, что возрастание скорости на порядок (то есть не менее чем в 10 раз) – это значительное преувеличение (так же, как и описание скорости развития генИИ и его влияния как сочетания шторма с цунами, о чем заявляет автор недавно вышедшей «Библии генеративного ИИ» [Fraley 2023]). Тем не менее заметного увеличения скорости разработки ИИ (с учетом подключения суперкомпьютеров ценой в миллиарды долларов) ожидать можно. Это должно давать лишь дополнительные стимулы к разумному ограничению разработок и требованиям прозрачных алгоритмов. Поэтому вполне вероятно, что через 10–15–20 лет возникнут такие продвинутые «LLM-приложения» (то есть большие языковые модели), вообразить которые мы сегодня не можем [Chat... 2023]. Не-

редко так и происходит в развитии инновационных технологий. Менее верится в другой прогноз одного из первопроходцев в области ИИ Терри Сейновски, что LLM станут окончательным информационным гаджетом людей, заменив все предыдущие. Социальные сети очень сильно потеснили сайты, но и те и другие пока существуют.

**Искусственный интеллект и образование.** Примерно через 10 лет образование будет основано на ИИ-преподавателях, считает тот же Т. Сейновски [Chat... 2023]. Однако мы не думаем, что такая замена произойдет в течение 10 лет. Сейчас больше опасности представляет возможность удаленного образования, которое резко понижает качество. Замена учителей и преподавателей на ИИ, конечно, будет происходить, но мы полагаем, что это гораздо более длительный процесс (как минимум несколько десятилетий) и такая замена едва ли может стать полной. Однако в некоторых узких сферах, прежде всего там, где люди готовы учиться самостоятельно (например, иностранным языкам), она вполне возможна, поскольку диалог с ИИ уже налажен на высоком уровне, поиск данных у большой языковой модели занимает короткое время, перевод с одного языка на другой налажен, так же как в целом фонетика и пр. Поэтому полностью игнорировать такую трансформацию было бы глупо, необходимо найти способы, которые бы позволили превратить ИИ в помощников в образовательном процессе, но не заменили людей квазипартнерами.

Возвращаясь к тому, о чем мы уже говорили выше, подчеркнем: нельзя допустить, чтобы те, кто контролирует ИИ, а также указанные инфорги, восторженно принимающие каждую инновацию, вытеснили самостоятельно и творчески мыслящих людей, гуманитариев. Для этого нужна, помимо ограничений в плане развития больших языковых моделей и сужения поля их обучения, также государственная политика, особенно в области образования, чтобы необходимые технические навыки в школе и институте любого профиля приобретались так же прочно, как навык письма, чтения и счета. Тогда и для гуманитария будет несложно вписаться в среду искусственного интеллекта, и это будет плодотворный симбиоз.

### ***3.2. Опасности, о которых говорят незаслуженно мало***

**ИИ и его влияние на социологические исследования.** Новые большие языковые модели активно используются для контент-

анализа социологических групп. Это важно для влияния на избирателей, покупателей, сочувствующих, оппонентов и т. п. Смысл – составить модели мышления определенных групп на основе анализа их текстов и другого контента, а затем постоянно мониторить изменения в количественном и качественном составе этих групп, чтобы эффективнее влиять на них. С учетом закрытости алгоритмов данных языковых моделей столь глубокое проникновение в человеческие настроения, привязанности, убеждения и т. п. становится все более опасным и неконтролируемым. А поскольку инструменты проникновения и влияния становятся все удобнее для манипуляторов, то возникает очень серьезная угроза приватности, политической свободе и свободному волеизъявлению. Фактически возможность манипулировать нашим выбором сильно возрастает.

Даже проекты, как будто преследующие благородную внешнюю цель, вроде проекта международной группы исследователей из Швейцарии, Австрии и Великобритании «Проектируемая демократия» (*Democracy by Design*), который направлен на развитие так называемой компьютерной дипломатии (*Computational Diplomacy*) и перспективы модернизации общества с помощью цифровых технологий, основанных на участии населения, имеет подозрительные опции. В частности, среди основных инструментальных направлений «проектируемой демократии»:

- формирование мнения людей;
- борьба с ложной информацией и дезинформацией;
- повсеместное использование коллективного разума для принятия решений;
- выстраивание системы голосований (ибо простое электронное голосование работает плохо и несет дополнительные риски манипуляций) [Helbing *et al.* 2023].

А мы прекрасно знаем, в каком направлении сейчас толкают демократию левые политики и политиканы, пытаясь «промывать мозги» избирателям. Но если до этого основным инструментом «промывки» была медиа, то теперь все большую роль будет играть ИИ, в частности большие языковые модели. И с учетом вышесказанного новые модели электронного голосования могут стать более изощренными в плане влияния на принятие решений.

**Мыслечтение и мыслепреступление.** Большие языковые модели активно обучают тому, чтобы читать чужие мысли без каких-либо вживляемых электродов. В частности, «Декодер смыслов» (генеративная модель типа ChatGPT) работает так. Сначала до 15 ча-

сов декодер учится сопоставлять паттерны мозговой активности лежащего в фМРТ-сканере и слушающего там подкасты человека с текстами этих подкастов. Потом, научившись сопоставлять паттерны и текст, декодер может генерировать тексты на основе любых паттернов мозговой активности человека (произвольных мыслей, воображаемых историй и т. д.). Результат является не дословной расшифровкой мыслей, а текстом, примерно соответствующим по смыслу услышанному человеком (при тренировке декодера) или подуманному им (при эксплуатации декодера).

Гипотеза авторов исследования «Клонирование мыслей: обучение мыслить в процессе действий, имитируя человеческое мышление» (Университет Британской Колумбии, Vector Institute и Canada CIFAR AI Chair) [Airhart 2023] заключается в том, что если вы обучаете модель действиям и соответствующим им мыслям, то она выучит правильные ассоциации между поведением и целями. Разумеется, все не так прямолинейно, процесс мышления намного сложнее теории, поэтому абсолютно не ясно, можно ли в принципе научить читать чужие мысли. Но, на наш взгляд, само движение в этом направлении является серьезной угрозой, а достижение даже небольших успехов в этом плане будет крайне опасным. Всем абсолютно понятно, кто и в каких целях будет этим пользоваться, а о приватности придется забыть совсем. Поэтому такого рода исследования необходимо остановить.

Разумеется, сейчас трудно судить, насколько обоснованы те или иные прогнозы. Очевидно, что большинство того, о чем пишут, не произойдет вовсе или не состоится даже близко к прогнозируемому объему. Однако то, о чем пишут меньше или не пишут вообще, может неожиданно быстро развиваться и распространиться. В любом случае развитие генеративных ИИ – это путь к усилению власти симбиоза правительств и крупнейших глобальных игроков, путь к окончательному попранию свобод и прав, при этом даже тех, на которые ранее никто не покушался (возьмем, к примеру, чтение мыслей, от которого недалеко и до конкретизации идеи мыслепреступления<sup>13</sup>). Это движение способно усилить глобали-

---

<sup>13</sup> Уже сегодня во многих странах, включая Россию, публичное воспроизведение тех или иных слов или выражений становится причиной общественного ostracизма, травли или даже уголовного преследования. А за сомнениями в до-

стов; однозначно возрастет контроль, инфильтрация идей, нужных заказчикам. Разумеется, появятся какие-то естественные сложности, преграды, барьеры и механизмы, уменьшающие власть генИИ (см. выше). Но при столь быстрой скорости развития и неясности перспектив и опасностей предварительное ограничение развития ИИ и его влияния должно стать императивом<sup>14</sup>. С другой стороны, идеологическая цензура не должна быть чрезмерной и самодовлеющей. Например, в Китае готовится закон, согласно которому незаконными будут признаваться применения любого генеративного ИИ, если для его обучения был использован обучающий набор данных, содержащий более 5 % «незаконного и вредного» контента. «Вредным контентом» в Китае считается вся информация, блокируемая «великим китайским файрволом» (то есть китайской цензурой в Интернете). Очевидно, что политические цели здесь стоят на первом месте. Таким образом, проблема выглядит очень сложной: как ограничить развитие генеративного ИИ, но не превратить его в инструмент усиления власти правительства?

Так или иначе, очевидно, что все сенсационные успехи ИИ больших языковых моделей (LLM) достигнуты за счет мощнейшего анализа плодов деятельности людей (то есть их действий) в виде оцифрованных данных. Это созданные людьми тексты самого разного содержания и объема: от научных до стихотворных; изображения, фотографии и видео; мелодии; различные действия, включая разговоры. На этой гигантской базе и учатся LLM. При этом их разработчики считают данную базу как бы ничейной, а фактически

---

стоверности каких-то исторических событий вроде холокоста (либо их масштаба) также может последовать криминальное обвинение. Но раз за слово (при свободе слова) могут посадить в тюрьму, почему бы не пойти дальше и не сажать за мысль?

<sup>14</sup> Где-то мы можем просто не успеть. В самом деле, единственная в истории атомная бомбардировка произошла, потому что никто еще не подозревал о последствиях такого оружия и шла война. А если бы бомбы создали на три-четыре месяца позже, подобной бомбардировки бы просто не было. То же и с COVID-19. Скорость распространения и навязывание образа действий не дали обществу контролировать применение вакцин. Теперь мы на протяжении многих лет будем узнавать о все новых тяжелых побочных эффектах вакцинирования. Скорость развития ИИ может привести к каким-то крайне нежелательным последствиям. Между тем глобалисты и власть имущие как раз и стремятся воспользоваться неопытностью и неорганизованностью общества, чтобы провести те или иные инновации, способные нанести серьезный вред.

своей, поскольку именно они имеют огромные преимущества в ее анализе и препарировании. Однако это не так. И пришло время (фактически оно пришло еще позавчера, но сегодня медлить уже нельзя) ограничить и регламентировать использование этой базы для целей обучения ИИ.

#### **4. Заключительные идеи**

##### **Контроль нужен не над технологией, но над ее заказчиками.**

В заключение мы хотели бы высказать вполне простую, неоднократно уже озвученную и понятную мысль, которая, однако, затерялась на фоне раздуваемой истерии по поводу будущего ИИ. Никакой нечеловеческий интеллект не является (и, думаем, никогда не будет являться) самостоятельным игроком. Какие бы чудеса в плане решения очень сложных интеллектуальных и творческих задач сегодня и завтра ни показывал нам ИИ, эти модели и программы всегда будут оставаться лишь инструментом, созданным и контролируемым людьми. Так же как с военными или биологическими (вспомним ковид) технологиями, опасность коренится не в могуществе нового генеративного ИИ, а в могуществе тех элит (сегодня глобальных), которые остаются без контроля со стороны общества. Эта бесконтрольность, защищаемая всеми формами подкупа, секретности, монополизма, сговора, мощи государственной машины и тайных служб, манипулирования, внушаемой идеологии, коррупции, якобы государственной или политической целесообразности, круговой поруки и прочими разнообразными способами властвовать и не отвечать за свои деяния, и является главной угрозой нашей приватности, нашим конституционным и иным правам, обществу и человечеству в целом.

И при таких всевластии, тотальном обмане и безнаказанности бесконтрольное развитие ИИ действительно становится очень опасным. Поэтому разумные ограничения развития ИИ, его обучения, бесконтрольного и бесплатного использования материалов, которые имеют правообладателей, требования прозрачности алгоритмов, а также демонополизации рынка и т. д. и т. п. становятся очень важной и крайне злободневной задачей. При правильном законодательстве и жестком контроле развитие ИИ станет достаточно безопасным и, несомненно, принесет много пользы.

Любая технология подконтрольна тем, кто ею управляет. По крайней мере, в первом назначении. Разнообразные другие влияния, разумеется, никто не просчитывает, как это всегда было при техническом прогрессе, внедрении компьютерных игр и т. п. [об опасностях будущих технологий см.: Гринин Л. Е., Гринин А. Л. 2023]. Однако с учетом быстрого развития ИИ, поскольку мы не можем просчитать более отдаленные последствия его прогресса, лучше сузить возможности искусственного интеллекта, сделать рамки разработки нового более жесткими. Проще потом их расширить, поскольку, если не ввести ограничения, загнать ИИ в жесткие рамки будет намного сложнее.

В любом случае должны быть выработаны правила безопасного использования ИИ, как давно сделано в отношении техники, химических средств, лекарств и мн. др.<sup>15</sup> Процесс, называемый по-английски *alignment* (грубо говоря, подгонка, балансировка целей и противоречий), направлен на то, чтобы поведение LLM было полезным для людей и не причиняло им вреда. Обычно это достигается путем настройки модели таким образом, чтобы усилить желаемое для нас поведение модели и ослабить нежелательное. Аналогичный процесс у людей называется воспитанием. Люди именно так воспитывают детей<sup>16</sup>.

Разумеется, можно согласиться с выводом о том, что есть фундаментальные ограничения такого «воспитания» больших языковых моделей [Wolf *et al.* 2023] и всегда возможны отклонения и инциденты. Но то же самое происходит и с техникой, которая иногда сильно подводит человека. Однако правила эксплуатации и безопасности делают такие техногенные катастрофы в целом редким и по большому счету терпимым явлением. Точно так же будет и с ИИ, если создать правила «безопасного генеративного искусственного интеллекта». И эти правила должны быть разработаны и внедрены повсеместно.

---

<sup>15</sup> Первые шаги в этом направлении уже делаются. Например, компания Anthropic (активный участник гонки ко все более умному ИИ) опубликовала свою «Политику ответственного масштабирования ИИ» (<https://www.anthropic.com/index/anthropic-responsible-scaling-policy>). Но здесь требуются очень серьезные изменения в законодательстве.

<sup>16</sup> Сергей Карелов [Малоизвестное... 2023] предлагает переводить *alignment* как «воспитание». Традиционные переводы этого термина применительно к ИИ – выравнивание, согласование, приведение в соответствие целей, предпочтений или этических принципов человека и машины – ему видятся менее точными. Нам нравится такая трактовка.

**Как далеко пойдет «игра человека с ИИ»?** Дэниэл Белл, автор теории постиндустриального общества, в числе измерений последнего называет и «изменение характера труда». Жизнь человека в рамках доиндустриального общества была, по его терминологии, «игрой между человеком и природой» (game against nature), то есть взаимодействием людей и природных сил и ресурсов. В индустриальном обществе природу заменяет искусственная среда (машины). Здесь налицо «игра человека с машиной» (man-machine relationship). В постиндустриальном обществе, в котором главной является сфера услуг, ведущей становится, по мысли Белла, «игра человека с человеком» (game between persons) [Bell 1973]. Д. Белл подчеркивает, что «люди должны учиться жить друг с другом» [Белл 1984: 20–21]. К сожалению или к счастью, но «игра человека с человеком» в настоящее время сильно сократила сферу своего действия. Зато новый тип – «игра человека с искусственным интеллектом» – быстро набирает обороты, в результате чего личное общение людей все более заменяется общением в Интернете. А теперь оно будет быстро замещаться общением с ИИ в виде Алисы, Алексы или других помощников. Это во многом неизбежно. Однако, перефразируя Д. Белла, «люди должны учиться и научиться жить с искусственным интеллектом». Но для этого нам необходимо обуздать стремления тех сил, которые хотят использовать ИИ для порабощения общества, для «перепрошивки» нашей психологии, для подмены свободы нашего выбора. Чтобы научиться жить в согласии с ИИ, нам нужно выработать для него четкие и жесткие правила и заставить его хозяев их соблюдать.

Английский философ XVI столетия Фрэнсис Бэкон говорил: «Деньги – очень хорошие слуги, но плохие господа». И эта мысль полностью подходит к ситуации с искусственным интеллектом.

\* \* \*

Мы движемся вперед все быстрее, но при этом всегда по неизведанному пути, на ощупь, очень слабо представляя последствия применения своих инноваций. И это не может не вызывать опасений. «Человек меняет жизнь на Земле с огромной скоростью, по большому счету даже не осознавая этого» (Филд 2015). Но уже давно пора начать отдавать себе отчет в последствиях каждого нового шага. И хотя у нас нет другого выбора, кроме как идти вперед, максимум осторожности, мудрости, благоразумия и даже некото-



рого смирения перед величием Вселенной и мира, глубокого уважения к наследию, оставленному нам миллиардами лет биологической эволюции, абсолютно необходим на этом пути. И тогда наши упорство, знания и (пусть еще слабая) способность к предвидению позволят благополучно дойти до новых вершин человеческого могущества и оставить потомков, способных его сохранить.

### *Литература*

Баранов Г. Nasdaq: рост на слухах – падение на фактах [Электронный ресурс] : Expert.ru. 2023. 12 сентября. URL: <https://expert.ru/2023/09/12/aktsii-ssha/> (дата обращения: 22.08.2023).

Белл Д. Постиндустриальное общество // «Американская модель»: с будущим в конфликте / под общ. ред. Г. Х. Шахназарова. М. : Прогресс, 1984. С. 16–24.

Бендерская Е. Н., Толстов А. А. Тенденции развития средств аппаратной поддержки нейровычислений // Научно-технические ведомости СПбГУ. 2013. № 3(174). С. 9–18.

Гаврилов А. В., Канглер В. М. Нейроморфные технологии: состояние и перспективы развития // Материалы VII Всероссийской научно-технической конференции «Робототехника и искусственный интеллект» РИ-2015. Железногорск, 2015. С. 148–154.

Гринин Л. Е., Гринин А. Л. Кибернетическая революция и шестой технологический уклад // Историческая психология и социология истории. 2015а. № 8(1). С. 172–197.

Гринин Л. Е., Гринин А. Л. От рубил до нанороботов. Мир на пути к эпохе самоуправляемых систем (история технологий и описание их будущего). М.: Моск. ред. изд-ва «Учитель», 2015б.

Гринин Л. Е., Гринин А. Л. Возможности и опасности технологий будущего // История и современность. 2023. № 1. С. 63–87.

Каспарьянц Д. О развитии базовых моделей ИИ. 2022 [Электронный ресурс]. URL: [https://rdc.grfc.ru/2022/07/ai\\_foundation\\_models\\_development/](https://rdc.grfc.ru/2022/07/ai_foundation_models_development/).

Малоизвестное интересное (авторский канал Сергея Карелова): [сайт]. URL: <https://t.me/theworldisnoteasy> (дата обращения: 26.08.2023).

Маркс К., Энгельс Ф. Манифест коммунистической партии / К. Маркс, Ф. Энгельс // Соч. 2-е изд. Т. 4. М.: Гос. изд-во полит. лит-ры, 1955. С. 419–459.

Разин А. В. Этика искусственного интеллекта // Философия и общество. 2019. № 1. С. 57–74. DOI: 10.30884/jfo/2019.01.04.

Скобцов Ю. А. Основы эволюционных вычислений. Донецк : ДонНТУ, 2008.

Филд Д. Идеальное генетическое знание [Электронный ресурс] : ИноСМИ. 2015. 11 сентября. URL: <http://inosmi.ru/world/20151005/230636715.html> (дата обращения: 31.08.2023).

Что такое генеративный искусственный интеллект? [Электронный ресурс] : Amazon. 2023. URL: <https://aws.amazon.com/ru/what-is/generative-ai/>.

Airhart M. 2023. Brain Activity Decoder Can Reveal Stories in People's Minds [Электронный ресурс]. URL: <https://cns.utexas.edu/news/podcast/brain-activity-decoder-can-reveal-stories-peoples-minds>.

Alharbi F. S. Dealing with Data Breaches Amidst Changes in Technology // International Journal of Computer Science and Security (IJCSS). 2020. Vol. 14(3). Pp. 108–115.

An Automated Pipeline for Understanding How the Brain is Wired [Электронный ресурс] : Allen Institute. 2020. February 10. URL: <https://alleninstitute.org/news/an-automated-pipeline-for-understanding-how-the-brain-is-wired/>.

Ashman H., Brailsford T., Cristea A. I., Sheng Q. Z., Stewart C., Toms E. G., Wade V. The Ethical and Social Implications of Personalization Technologies for E-learning // Information & Management. 2014. Vol. 51(6). Pp. 819–832. DOI: 10.1016/j.im.2014.04.003.

Bell D. The Coming of Post-Industrial Society: A Ventury in Social Forecasting. New York : Basic Books, Inc, 1973.

Biswas S. ChatGPT and the Future of Medical Writing // Radiology. 2023. February 2. 223312. DOI: 10.1148/radiol.223312.

Brammer S., Branicki L., Linnenluecke M. COVID-19, Societalization and the Future of Business in Society // Science of the Total Environment. 2020. Vol. 34(4). Pp. 2–7. DOI: 10.5465/amp.2019.0053.

Cecere, G., Le Guel, F., Soulié, N. Perceived Internet Privacy Concerns on Social Networks in Europe // Technological Forecasting and Social Change. 2015. Vol. 96. Pp. 277–287. DOI: 10.1016/j.techfore.2015.01.021.

Cepelewicz J. The U.S. Government Launches a \$100-million “Apollo Project of the Brain” [Электронный ресурс] : Scientific American. 2016. March 8. URL: <https://www.scientificamerican.com/article/the-u-s-government-launches-a-100-million-apollo-project-of-the-brain/> (дата обращения: 29.08.2023).

Chat GPT and the Talking Dog with Dr. Terry Sejnowski. 2023 [Электронный ресурс]. URL: <https://www.youtube.com/watch?v=dZOEXNlrZLI> (дата обращения: 30.08.2023).

Fraleay A. AI Bible: [5 in 1] The Most Updated and Complete Guide | From Understanding the Basics to Delving into GANs, NLP, Prompts, Deep Learning, and Ethics of AI. Kerala : AlgoRay Publishing, 2023.

Hawkins J., Blakeslee S. On Intelligence. New York : Owl Books, 2004.

Helbing D., Mahajan S., Hänggli Fricker R., Musso A., Hausladen C. I., Carissimo C., Carpentras D., Stockinger E., Argota Sanchez-Vaquerizo J., Yang J. C., Ballandies M. C., Korecki M., Dubey R. K., Pournaras E. Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society [Электронный ресурс] : Journal of Computational Science. 2023. Vol. 71. URL: <https://www.sciencedirect.com/science/article/pii/S1877750323001217?via%3Dihub>.

Gates B. The Age of AI has Begun: Artificial Intelligence is as Revolutionary as Mobile Phones and the Internet. 2023 [Электронный ресурс]. URL: <https://www.gatesnotes.com/The-Age-of-AI-Has-Begun> (дата обращения: 28.08.2023).

Gödel, Escher, Bach. Author Doug Hofstadter on the State of AI Today [Электронный ресурс] : YouTube. 2023. URL: <https://www.youtube.com/watch?v=lfXxzAVtdpU> (дата обращения: 23.08.2023).

Grinin L., Grinin A. The Cybernetic Revolution and the Forthcoming Epoch of Self-Regulating Systems. Moscow : Uchitel, 2016.

Grinin L., Grinin A., Korotayev A. Forthcoming Kondratieff Wave, Cybernetic Revolution, and Global Ageing // Technological Forecasting and Social Change. 2017a. Vol. 115. Pp. 52–68. DOI: 10.1016/j.techfore.2016.09.017.

Grinin L., Grinin A., Korotayev A. The MANBRIC-Technologies in the Forthcoming Technological Revolution // Industry 4.0 – Entrepreneurship and Structural Change in the New Digital Landscape: What is Coming on Along with the Fourth Industrial Revolution / ed. by T. Devezas, J. Leitão, A. Sarygulov. N. p. : Springer, 2017b. Pp. 243–261. DOI: 10.1007/978-3-319-49604-7\_13.

Grinin L., Grinin A., Korotayev A. A Quantitative Analysis of Worldwide Long-Term Technology Growth: From 40,000 BCE to the Early 22<sup>nd</sup> Century // Technological Forecasting and Social Change. 2020. Vol. 155. Pp. 1–15.

Grinin L., Grinin A., Korotayev A. Does COVID-19 Accelerate the Cybernetic Revolution and Transition from E-government to E-state? // Kondratieff Waves: Processes, Cycles, Triggers, and Technological Paradigms / ed. by L. E. Grinin, A. V. Korotayev. Volgograd: Uchitel, 2021. Pp. 95–125.

IBM Institute for Business Value. Augmented Work for an Automated, AI-driven World [Электронный ресурс]. URL: <https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/augmented-workforce>.

Lund B. D., Ting Wang. Chatting about ChatGPT: How May AI and GPT Impact Academia and Libraries? // Library Hi Tech News. 2023. Vol. 40. DOI: 10.1108/LHTN-01-2023-0009.

Moustaka V., Theodosiou Z., Vakali A., Kounoudes A., Anthopoulos L. G. Enhancing Social Networking in Smart Cities: Privacy and Security Border-

lines // *Technological Forecasting and Social Change*. 2019. Vol. 142. Pp. 285–300. DOI: 10.1016/j.techfore.2018.10.026.

Ott T. OpenAI Lawsuit: US Authors Allege ChatGPT Copyright Theft. 2023 [Электронный ресурс]. URL: <https://www.dw.com/en/openai-lawsuit-us-authors-allege-chatgpt-copyright-theft/a-66895907?maca=en-rss-en-all-1573-rdf>.

Rudnitsky J. ChatGPT to Fuel \$1.3 Trillion AI Market by 2032, New Report Says [Электронный ресурс] : Bloomberg. 2023. June 1. URL: <https://www.bloomberg.com/news/articles/2023-06-01/chatgpt-to-fuel-1-3-trillion-ai-market-by-2032-bi-report-says> (дата обращения: 30.08.2023).

Schmidhuber J. Deep Learning in Neural Networks: An Overview. Technical Report IDSIA-03-14. 2014 [Электронный ресурс]. URL: <http://arxiv.org/pdf/1404.7828.pdf> (дата обращения: 29.08.2023).

Schwartz P. M. Internet Privacy and the State // *Connecticut Law Review*. 1999. Vol. 32. Pp. 815–829.

Singularity Hypotheses: A Scientific and Philosophical Assessment / ed. by A. H. Eden, J. H. Moor, J. H. Søraker, E. Steinhardt. Berlin : Springer, 2012. DOI: 10.1007/978-3-642-32560-1.

Solove D. J. *Understanding Privacy*. Harvard : Harvard University Press, 2008.

Westin A. F. Science, Privacy, and Freedom: Issues and Proposals for the 1970's. Part I. The Current Impact of Surveillance on Privacy // *Columbia Law Review*. 1966. Vol. 66(6). Pp. 1003–1050.

Wolf Y., Wies N., Avnery O., Levine Y., Shashua A. 2023. Fundamental Limitations of Alignment in Large Language Models. 2023 [Электронный ресурс]. URL: <https://arxiv.org/abs/2304.11082>.