

СЕРГЕЙ ПЕТРОВИЧ ПРОХОРОВ

кандидат физико-математических наук,
ведущий научный сотрудник
Института истории естествознания и техники
им. С.И. Вавилова Российской академии наук,
Москва, Россия;
e-mail: sergei.prokhorov@gmail.com



Сингулярность и искусственный интеллект

УДК: 004.8

DOI: 10.24412/2079-0910-2024-4-23-38

В 1950 г. Алан Тьюринг опубликовал в британском академическом журнале *MIND* статью “Computing Machinery and Intelligence”, в которой сформулировал вопрос «Могут ли машины мыслить?». В этой работе он, в частности, рассмотрел и отверг возражения против способности машины поступать подобно человеку. Тьюринг предсказал, что будущее за машинами, которые будут способны к самообучению.

Сегодня мы видим, что его пророчество сбылось. Развитие вычислительной техники, параллельных вычислений, облачных вычислений, технологии нейросетей, глубокого обучения, а также больших языковых моделей привело к тому, что мы находимся на пороге создания машинного суперинтеллекта, который превзойдет способности человека. О рисках, связанных с развитием универсального искусственного интеллекта (ИИ), предупреждает большой круг известных ученых и общественных деятелей, в частности, Джеффри Хинтон, лауреат Нобелевской премии 2024 г. «за работы в области искусственных нейросетей и применения их для машинного обучения». Ряд ученых считают, что риски преувеличены, что сингулярность — момент, когда способности ИИ превзойдут интеллектуальные способности человека, — никогда не наступит. Но в любом случае, по мнению экспертов ЮНЕСКО, «быстрый технологический прогресс в области искусственного интеллекта бросает вызов представлениям о том, что значит быть человеком».

Ключевые слова: А. Тьюринг, искусственный интеллект, большие языковые модели, глубокое обучение, сингулярность.

Последнее десятилетие ознаменовалось бурным ростом исследований, связанных с развитием применения технологии искусственного интеллекта (ИИ). В настоящее время ИИ применяется во многих областях человеческой деятельности. Он существенно улучшает качество услуг, производительность труда, положительно влияет на экономические показатели многих отраслей. Однако наряду с открывающимися широкими перспективами возникают и серьезные риски и угрозы при-

менения ИИ для человечества. По этой причине ООН, ЮНЕСКО и другие авторитетные международные организации приняли ряд постановлений и рекомендаций, призванных уменьшить потенциальную угрозу от внедрения ИИ.

В Российской Федерации документом, регламентирующим работы в области ИИ, является Национальная стратегия развития искусственного интеллекта на период до 2030 года¹. Этот документ определяет искусственный интеллект как «комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их». Подчеркивается, что «комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе, в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений».

Современное применение технологии ИИ в значительной степени связано с ростом мощности вычислительных ресурсов, что в совокупности с развитием алгоритмов машинного обучения позволило обеспечить резкий рост машинной обработки структурированных и неструктурированных данных в последние два десятилетия.

Алан Тьюринг был первым, кто обратился к проблеме создания ИИ. В 1950 г. он опубликовал в британском академическом журнале *MIND* статью “Computing Machinery and Intelligence”, в которой сформулировал вопрос «Могут ли машины мыслить?» В этой работе он, в частности, рассмотрел и отверг возражения против способности машины поступать подобно человеку. Тьюринг предсказал, что будущее за машинами, которые будут способны к самообучению [Turing, 1950].

Искусственный интеллект как академическая дисциплина возник в 1956 г. Сам термин был предложен Джоном Маккарти (John McCarthy), одним из организаторов семинара по вопросам искусственного интеллекта, проведенного в Дартмутском колледже летом 1956 г. Дартмутский семинар не принес каких-либо крупных открытий, но по его итогам были сформулированы основные положения нового направления в науке.

В СССР родоначальником ИИ, без сомнения, является выдающийся советский математик Александр Семенович Кронрод. В 1949 г. он был приглашен в Институт теоретической и экспериментальной физики (ИТЭФ), где возглавил математический отдел. Поначалу расчеты велись на электрических арифмометрах «Мерседес». Только в 1955 г. А.С. Кронрод и его сотрудники получили возможность работать на одной из первых советских электронных вычислительных машин (ЭВМ) М-2, созданной в Энергетическом институте Академии наук.

Незадолго до этого, в пятом номере журнала «Вопросы философии» за 1953 г., за анонимной подписью «Материалист» появилась статья «Кому служит кибернетика». В этой статье кибернетика объявлялась «архиреакционной теорией». Ее последователи получали ярлык «ученых лакеев капитализма», так как «утверждают, что

¹ Указ Президента Российской Федерации № 490 от 10 октября 2019 г. «О развитии искусственного интеллекта в Российской Федерации. Национальная стратегия развития искусственного интеллекта на период до 2030 года». Режим доступа: <http://www.kremlin.ru/acts/bank/44731> (дата обращения: 24.10.2024).

деятельность вычислительных машин дает ключ к познанию как биологических, так и социальных явлений» [*Материалист*, 1953]. Фактически применение вычислительных машин для каких-либо целей, кроме расчетов, было запрещено. Отметим, что к моменту публикации этой статьи пальцев одной руки хватило бы, чтобы пересчитать все ЭВМ, созданные в Советском Союзе. Все они были загружены работами по атомному проекту и другим темам, связанным с оборонной тематикой, так что статья на текущие работы никак повлиять не могла. Но это был выстрел в будущее!

К счастью, у ЭВМ нашлись весьма влиятельные заступники. В том числе академик Леонид Сергеевич Соболев, один из величайших математиков XX в., который имел очень высокий авторитет в руководстве страны. Было разрешено провести публичную дискуссию в университетах и научных центрах, чтобы обсудить и альтернативную точку зрения. Результатом стала публикация в 1955 г. в журнале «Вопросы философии» еще одной статьи под названием «Основные черты кибернетики», где авторы — С.Л. Соболев, А.И. Китов и А.А. Ляпунов — объяснили, что основы кибернетики ни в коем случае не подрывают основы марксистско-материалистической идеологии [*Соболев и др.*, 1955]. Эта публикация стала сигналом, что отныне ЭВМ разрешено применять не только для расчетов оборонного характера и бухгалтерии, но и для моделирования широкого круга процессов.

Так что к моменту появления М-2 в ИТЭФ запреты были сняты. Начав заниматься программированием, А.С. Кронрод быстро пришел к мысли, что использовать ЭВМ только для программирования вычислений не слишком увлекательная идея. Гораздо интереснее — попытаться научить машину думать! Так возник первый в СССР семинар по «эвристическому программированию». Это то, что в настоящее время называется «искусственным интеллектом».

А.С. Кронрод предложил выбрать эталонную задачу, решение которой дало бы представление об уровне, достигнутом участниками семинара в области эвристического программирования. Первая задача, которая была выбрана и запрограммирована, — карточная игра в «подкидного дурака». Это была игра, не имеющая разработанной теории, но позволяющая простое описание состояния позиции. Программа была написана и хорошо играла в ситуации с «неполной информацией», то есть когда колода еще не была разобрана. Когда дело доходило до игры «в открытую», мощности машины не хватало для перебора вариантов.

Следующей вехой стала игра в шахматы. С Джоном Маккарти (автором термина «искусственный интеллект») договорились провести матч между шахматной программой ИТЭФ и программой, созданной под руководством Дж. Маккарти в Стэнфордском университете. Матч состоялся в 1967 г. Времени для счета выделялось мало, так как плановые расчеты, естественно, имели приоритет. Неудивительно, что матч из четырех партий растянулся почти на год.

Этот матч имел чисто научное значение. Политики вокруг него не было. Победа не была принципиальной — главное, чтобы дальше в исследованиях можно было опираться на результат и наука двигалась вперед. Программа ИТЭФ выиграла со счетом 3:1. Это позволило команде А.С. Кронрода получить разрешение работать в этом направлении дальше. Американцы извлекли пользу даже еще большую. Они сообщили своему начальству, что есть область, где США отстают от СССР, и им позарез нужно дополнительное финансирование, чтобы это исправить. Финансирование было получено.

В 1974 г. состоялся первый чемпионат мира по шахматам среди компьютерных программ. Советская программа «Каисса» на чемпионате мира заняла первое место по очкам, а американская программа *Chess 4.0* — второе [Ландис, Яглом, 2001; *Russian Math*, 2001].

Первые 20 лет развития ИИ можно условно назвать «романтическим периодом». Все были полны оптимизма, что с приходом ЭВМ можно будет быстро формализовать многие задачи и добиться впечатляющих прорывных результатов в познании мира. А знаменитая фраза Архимеда «Дайте мне точку опоры, и я переверну Землю» воспринималась в несколько измененном виде как «Дайте мне компьютер, и я изменю мир». Но впечатляющих успехов достигнуто не было, и наступила «зима ИИ», которая продолжалась, с кратковременными оттепелями, более 30 лет. Для этого периода характерно снижение как профессионального и общественного интереса к ИИ, так и резкое сокращение финансирования. Бывали периоды оттепели, когда интерес к ИИ на какое-то время возрастал (экспертные системы, распознавание образов, робототехника и т. д.), после которых ИИ вновь уходил на периферию научного и общественного интереса.

Возрождение общественного интереса к ИИ первоначально было связано с успехом его применения в машинном переводе и робототехнике, с интернетом вещей, развитием сферы интеллектуального анализа данных. Хотя во многих случаях термин «искусственный интеллект» не использовался, но всегда подразумевалось, что речь идет о частичной замене человеческой деятельности вычислительными системами в анализе данных, управлении и других интеллектуальных сферах деятельности человека.

Но наиболее сильное влияние на развитие ИИ оказала совокупность таких факторов, как значительный рост вычислительных ресурсов и сетей передачи данных, в результате чего стало возможно ведение параллельных вычислений; техническое совершенствование средств хранения больших объемов данных; создание больших банков данных; развитие облачных вычислений; развитие глубоких нейросетей; создание на их основе больших лингвистических моделей; технологии глубокого обучения. Все это вместе создало сильный кумулятивный эффект.

Общественный интерес и финансирование работ по ИИ значительно возросли после 2012 г., когда технология глубокого обучения ИИ показала весьма впечатляющие результаты по отношению к существующим. Интерес к ИИ и финансирование работ в этой области еще больше усилились после 2017 г. с появлением архитектуры *Transformer*. Возникший бум и широкое использование ИИ выявило ряд проблем, связанных с возможными непредсказуемыми последствиями его применения, и привело к пониманию, что дальнейшее развитие технологии ставит актуальные вопросы регулирования применения ИИ в общественно значимых сферах деятельности человека (в здравоохранении, образовании и др.). Это касается также и этических вопросов общения с ИИ. То есть речь идет о принятии правил применимости.

Об актуальности и значимости этой проблемы говорит тот факт, что она стала предметом обсуждения в таких авторитетных организациях, как ООН и ЮНЕСКО, которые приняли ряд документов — рекомендаций по безопасному применению ИИ, учитывающих этические проблемы общения человека и ИИ.

21 марта 2024 г. Генеральная Ассамблея ООН единогласно приняла резолюцию «Использование возможностей безопасных, защищенных и надежных систем искусственного интеллекта для устойчивого развития», в которой подчеркивалось,

что риски, связанные с ИИ, касаются всех — всего мира, всех стран. В документе предлагалось внедрять меры для соблюдения прав интеллектуальной собственности, обеспечения конфиденциальности и защиты персональных данных, а также обеспечения прозрачности, предсказуемости, надежности работы систем искусственного интеллекта на протяжении всего их жизненного цикла. Основной посыл документа: искусственный интеллект нужно регулировать, не препятствуя инновациям².

В развитие принятой резолюции Консультативный орган высокого уровня по искусственному интеллекту, созданный главой ООН, представил 19 сентября 2024 г. глобальный план действий по устранению рисков, связанных с ИИ, в итоговом докладе «Управление искусственным интеллектом в интересах человечества»³.

В докладе предлагалось организовать в рамках ООН межправительственный консультативный совет по вопросам развития и управления ИИ, который будет содействовать разработке и применению международных стандартов в области ИИ. Предлагалось создать независимую международную научную группу по ИИ, которая будет выпускать исследования и отчеты о возможностях ИИ и возникающих проблемах, способствовать продвижению внедрения ИИ-технологий в развивающихся странах. В докладе отмечалось, что ИИ несет различные риски для мира и безопасности, в том числе связанные с энергопотреблением постоянно растущего числа ИИ-систем. Поэтому разработку, внедрение и использование таких технологий нельзя оставлять только на усмотрение небольшого числа частных многонациональных компаний, обладающих значительными финансовыми ресурсами и профессионалами очень высокого уровня.

Генеральная конференция Организации Объединенных Наций по вопросам образования, науки и культуры (ЮНЕСКО) в ноябре 2021 г. приняла «Соглашение об этических аспектах искусственного интеллекта» — первый глобальный нормативный документ в этой области⁴. Принятый документ ставил своей целью уменьшить риски, связанные с внедрением технологии искусственного интеллекта, возможными негативными последствиями влияния ИИ на жизнь человека, в том числе на его интеллектуальную деятельность, обусловленную появлением новых возможностей, позволяющих ИИ влиять на процессы мышления, взаимодействия и принятия человеком решений, а также на сферу образования, гуманитарные, социальные и естественные науки, культуру и информационно-коммуникационные технологии.

С момента принятия Рекомендации ЮНЕСКО развивает сотрудничество с частным сектором, что привело к созданию Делового совета по этике ИИ, сопредседателями которого являются *Microsoft* и *Telefonica*. В феврале 2024 г. ЮНЕСКО совместно с восемью глобальными технологическими компаниями (*GSMA*, *INNIT*,

² Использование возможностей безопасных, защищенных и надежных систем искусственного интеллекта для устойчивого развития. Организация Объединенных Наций. Режим доступа: <https://documents.un.org/doc/undoc/ltd/n24/065/94/pdf/n2406594.pdf> (дата обращения: 24.10.2024).

³ В ООН представили план по управлению искусственным интеллектом. Организация Объединенных Наций. Режим доступа: <https://news.un.org/ru/story/2024/09/1456466> (дата обращения: 24.10.2024).

⁴ Государства — члены ЮНЕСКО принимают первое глобальное соглашение по этическим аспектам искусственного интеллекта. ЮНЕСКО. Режим доступа: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (дата обращения: 24.10.2024).

Lenovo Group, LG AI Research, Mastercard, Microsoft, Salesforce и Telefonica) подписания по этическим аспектам ИИ. Компании будут учитывать ценности и принципы Рекомендации ЮНЕСКО по этическим аспектам ИИ при разработке и внедрении систем ИИ. Это первый случай, когда частные компании сотрудничают с ООН (ЮНЕСКО) в этой области⁵.

По сути единогласное принятие резолюций ООН и ЮНЕСКО по ИИ свидетельствует, что правительства всех стран, всех континентов решительно перешли от вопроса о том, следует ли регулировать ИИ, к неотложному вопросу о том, как это сделать.

В РФ научные исследования и практические разработки в сфере ИИ регламентируются указом Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» и утвержденной тем же указом Национальной стратегией развития искусственного интеллекта на период до 2030 года. Указом Президента РФ от 15 февраля 2024 г. № 124 «О внесении изменений в Указ Президента Российской Федерации от 10 октября 2019 г. № 490 “О развитии искусственного интеллекта в Российской Федерации” и в Национальную стратегию, утвержденную этим указом» были внесены изменения, уточняющие положения, цели и задачи, которые должны быть достигнуты к 2030 г.

Первым откликом на указ Президента стало создание Альянса в сфере искусственного интеллекта, инициаторами которого выступили крупнейшие госкорпорации и компании РФ. Главный результат деятельности Альянса — разработка Кодекса этики в сфере ИИ. Этот документ носит рекомендательный характер. В нем перечисляются общие этические принципы и стандарты поведения, которыми следует руководствоваться участникам отношений в сфере искусственного интеллекта⁶.

В настоящее время Кодекс этики искусственного интеллекта подписали уже более 820 организаций, в том числе 27 зарубежных участников из 20 стран, которые выразили готовность руководствоваться принципами и рекомендациями Кодекса. Среди подписантов все федеральные ведомства, а также университеты, фирмы и частные организации.

Различаются слабый и сильный ИИ. Под слабым искусственным интеллектом понимается алгоритм, способный решать узкоспециализированные проблемы, возможно, даже лучше человека. Примерами слабого ИИ являются беспилотные автомобили, роботизированные системы, системы диагностики в медицинской сфере. В принципе для систем узкого интеллекта могут быть разработаны стандарты на его применение, которые тем не менее должны учитывать предметную область, в которой работает ИИ [Kuleshov, Prokhorov, 2019].

Такие стандарты уже создаются. Например, в РФ для клинической медицины разработано и введено в действие уже 12 стандартов на применение технологий ИИ. Однако все они касаются аспектов применения оборудования, но не лечения. Это связано с тем, что врачу не всегда понятно, на основании каких данных ИИ

⁵ Этика ИИ: 8 глобальных технологических компаний обязуются выполнять Рекомендацию ЮНЕСКО. ЮНЕСКО. Режим доступа: <https://www.unesco.org/ru/articles/etika-ii-8-globalnykh-tekhnologicheskikh-kompaniy-obyazuyutsya-vypolnyat-rekomendaciyu-yunesko> (дата обращения: 28.10.2024).

⁶ Альянс в сфере искусственного интеллекта. Кодекс этики в сфере ИИ. Режим доступа: <https://ethics.a-ai.ru> (дата обращения: 28.10.2024).

ставит свои диагнозы и предложения по лечению. В таком случае кто будет нести ответственность за применение выводов и рекомендаций ИИ? Врач, разработчики программ, специалисты, проводившие обучение ИИ на некачественном наборе датасетов? Ответа нет.

Эта проблема — отсутствие понимания того, как ИИ принимает решения или выдает рекомендации, — указана как одна из ключевых в Национальной стратегии развития. «Отсутствие понимания того, как искусственный интеллект достигает результатов, является одной из причин низкого уровня доверия к современным технологиям искусственного интеллекта и может стать препятствием для их развития»⁷. Это справедливо и по отношению к другим системам применения слабого ИИ. В частности, к ответственности за ДТП, которое произошло при участии беспилотного автомобиля. Вопрос о юридической ответственности является одним из основных тормозов широкого внедрения ИИ.

Еще одна проблема связана с тем, что слабый ИИ уязвим для предвзятостей и ошибок. Он извлекает информацию из ограниченного набора данных и решает только те задачи, на которые запрограммирован. Если данные искажены, слабый ИИ может выдавать необъективный предвзятый результат.

Появление концепции универсального (сильного) искусственного интеллекта напрямую связано с развитием облачных вычислений, технологии нейросетей и с появлением больших языковых моделей.

В 2017 г. восемь сотрудников *Google*, работавших над проблемой машинного перевода, опубликовали статью “Attention Is All You Need”⁸, которая оказала сильнейшее влияние на развитие работ в области ИИ [*Vaswani et al.*, 2027].

В статье была представлена новая архитектура глубокого обучения, известная ныне как «Трансформер», основанная на методе машинного обучения, который определяет относительную важность каждого компонента в последовательности по отношению к другим компонентам этой последовательности. Это дало сильнейший импульс для развития работ по ИИ. По состоянию на 2024 г. статья была процитирована более 100 000 раз. «Трансформер» стал основной архитектурой больших языковых моделей, которые используются в современных методах глубокого машинного обучения.

Применение технологии трансформера в больших языковых моделях позволяет обеспечить как более адекватный перевод семантики языковых конструкций во внутреннее машинное представление, так и генерацию выражений на естественном языке. Технологии глубокого (глубинного) обучения формируют в процессе обучения слои на нескольких уровнях представлений, которые соответствуют различным уровням абстракции. Слои образуют иерархию понятий, причем обучение может быть как контролируемым, так и неконтролируемым. В этом случае система сама формирует абстракции и их иерархию. Все это вместе приближает нас к моменту, когда способности ИИ воспринимать исходный текст, получать из него знание, абстрагировать его, анализировать, генерировать гипотезы и выносить суждения

⁷ Указ Президента Российской Федерации № 490 от 10 октября 2019 г. «О развитии искусственного интеллекта в Российской Федерации. Национальная стратегия развития искусственного интеллекта на период до 2030 года». Режим доступа: <http://www.kremlin.ru/acts/bank/44731> (дата обращения: 24.10.2024).

⁸ Название созвучно знаменитой песне «Битлз» “All You Need Is Love”.

будут сравнимы с уровнем человека, а со временем и превзойдут его. Этот момент называется сингулярностью.

Сам термин применительно к гипотетической ситуации, когда ИИ приведет к появлению мощного сверхинтеллекта, качественно намного превосходящего весь человеческий интеллект, впервые был предложен американским писателем-фантастом Вернором Винджем (Vernor Vinge) в 1983 г. Другим значительным фактором, способствовавшим более широкому распространению этого понятия, стала книга Рэя Курцвейла (Ray Kurzweil) 2005 г. «Сингулярность близка» (“The Singularity Is Near”), в которой он предсказывал достижение сингулярности к 2045 г.

Но сама идея о том, что развитие компьютерных технологий способно однажды привести к ситуации, когда компьютер сможет успешно заменять человека в некоторых областях деятельности, обсуждалась с момента создания первого компьютера. Начало этому положила статья Алана Тьюринга. Однако в большей степени это было уделом писателей-фантастов.

Качественный скачок произошел с созданием больших языковых моделей на базе трансформеров, которые позволили выявлять семантику языковых выражений, учитывая контекст. Это был революционный, переломный момент в развитии ИИ.

В настоящее время существует около 60 больших языковых моделей. Некоторые из них используются в исследовательских целях. О них мало что известно. Но к большинству из них можно получить доступ, подписав соглашение о правилах использования.

Появление чат-ботов является, пожалуй, самым ярким примером внедрения ИИ в нашу повседневную жизнь. Наиболее известен из них *ChatGPT* — генеративный чат-бот, разработанный *OpenAI* и запущенный в 2022 г. *ChatGPT* может генерировать разговорные ответы, похожие на человеческие, и позволяет пользователям уточнять и направлять разговор в соответствии с желаемой длиной, форматом, стилем, уровнем детализации и языком. Его появление привело к резкому повышению общественного внимания к области ИИ и, как следствие, к новым значительным инвестициям в ИИ. К январю 2023 г. *ChatGPT* стал самым быстрорастущим потребительским программным приложением в истории, набрав более 100 млн пользователей за два месяца.

В России лидерами разработок универсальных систем ИИ являются «Сбербанк» и «Яндекс». На базе *YandexGPT* работает хорошо известный виртуальный помощник «Алиса».

С разработкой систем ИИ на базе больших языковых моделей вновь обострился вопрос о том, насколько можно доверять ИИ. Например — при найме персонала. Такой вопрос стал перед руководством *IBM* и *Amazon*, когда они предприняли попытку использовать ИИ для найма сотрудников. Выяснилось, что ИИ в большей степени отдавал предпочтение мужчинам, чем женщинам. Это произошло потому, что система была обучена с использованием данных, собранных за 10-летний период, когда резюме поступали в основном от кандидатов-мужчин. Как следствие, ведущие компании стали привлекать экспертов для выявления предубеждения и рисков дискриминации, которые могли возникнуть в результате обучения ИИ на некачественном материале. Одним из таких экспертов стал Блейк Лемуан (Blake Lemoine), который был привлечен *Google* для тестирования *LaMDA* (Language Model for Dialogue Applications) перед тем, как сдать ее в эксплуатацию. В его задачу входило

ло тестирование *LaMDA* с помощью чат-бота, который был создан, чтобы увидеть, содержит ли *LaMDA* предвзятость в отношении сексуальной ориентации, пола, религии, политической позиции и этнической принадлежности.

Лемуан провел несколько часов в разговоре с чат-ботом, сборную стенограмму которого опубликовал в своем блоге [Lemoine, 2022]. Из опубликованной Лемуаном стенограммы диалога с *LaMDA* видно, что речь идет в основном об эмоциональной и чувственной стороне жизни человека и о том, как это может интерпретироваться *LaMDA*. Проанализировав свои диалоги с *LaMDA*, Лемуан пришел к выводу, что ИИ достиг уровня сознания человека, а эмоции выражает надежно и в правильном контексте. Опубликовав стенограмму интервью, он обратился в Конгресс, чтобы привлечь внимание к тому факту, что *LaMDA* нуждается в адвокате, так как является личностью.

Просто отмахнуться от публично высказанного мнения Лемуана как от мнения некомпетентного специалиста было невозможно. Во-первых, Лемуан был опытным и квалифицированным специалистом в области ИИ. Во-вторых, он происходил из консервативной христианской семьи и сам был проповедником, так что хорошо знал, что «Бог создал человека по своему образу и подобию». И вопрос о возможности создания человеком субъекта, обладающего разумом, был рассмотрен им тщательным образом.

Была создана комиссия, которая пришла к выводу: ничего особенного в общении *LaMDA* с человеком нет. Все ответы — это не более чем тексты, сгенерированные большой языковой моделью на основании ранее представленной информации. Сразу последовал риторический вопрос: а разве это не то же самое, что делает человек?

В июле 2022 г. Лемуан был уволен из *Google* за нарушение правил работы с корпоративными документами. Споры, вызванные инцидентом, побудили руководителей *Google* принять решение не выпускать *LaMDA* в режим общего пользования. История имела большой резонанс и вновь вызвала сильный общественный интерес к сингулярности.

Интересный факт: Рэй Курцвейл (Ray Kurzweil) — один самых ярых сторонников скорого наступления момента сингулярности — был в *Google* руководителем лаборатории, в которой работал Блейк Лемуан. Два года спустя, в июне 2024 г., он дал большое интервью. К этому моменту он написал еще одну книгу, «Сингулярность уже близко» (“The Singularity Is Nearer”), в которой утверждал, что еще до конца 2020-х гг. искусственный интеллект достигнет человеческого уровня. Ему был задан вопрос, как он относится к высказываниям Лемуана. Ответ был: «Я никогда не говорил, что он неправ. Очень скоро мы дойдем до момента, когда системы универсального ИИ будут действовать и реагировать так же, как люди. И если вы скажете, что они не обладают сознанием, вам придется сказать, что люди не обладают сознанием»⁹.

О рисках, связанных с развитием универсального ИИ, говорил и Джеффри Хинтон (Geoffrey Hinton), лауреат Нобелевской премии 2024 г. «за работы в области искусственных нейросетей и применения их для машинного обучения». С 2013 по 2023 г. он работал в *Google* (Google Brain). В мае 2023 г. он ушел из *Google*, чтобы

⁹ Levy S. If Ray Kurzweil Is Right (Again), You'll Meet His Immortal Soul in the Cloud. Available at: <https://www.wired.com/story/big-interview-ray-kurzweil/> (date accessed: 28.10.2024).

иметь возможность «свободно высказываться о рисках искусственного интеллекта». После получения Нобелевской премии он призвал к срочным исследованиям в области безопасности ИИ, чтобы выяснить, как управлять системами ИИ, которые умнее человека.

Мнения среди экспертов относительно момента достижения сингулярности неоднозначны, но большинство из них считают, что это произойдет в ближайшем будущем. К ним относятся такие выдающиеся личности, как Стивен Хокинг, Билл Гейтс и Илон Маск, указывающие на риск от возможного появления сверхразумного ИИ. Есть и авторитетные критики, которые утверждают, что компьютеры или машины не могут достичь человеческого интеллекта. В качестве доказательства своего довода они ссылаются на то, что чтение книг само по себе еще не гарантирует, что читатель их имеет выдающиеся способности. Но, как справедливо отметила Татьяна Владимировна Черниговская, — впечатляет не количество информации, закачанное в ИИ, а способность ИИ быстро и качественно учиться. Если это качество есть, то человек ИИ уже не нужен¹⁰.

В феврале 2023 г. *The New York Times* опубликовала стенограмму диалога своего сотрудника Кевина Руца (Kevin Roose) с чат-ботом *Microsoft Bing* [Roose, 2023]. Это была новейшая версия *Bing*, доступная только небольшой группе тестировщиков, на базе технологии ИИ *OpenAI*, производителя *ChatGPT*. Самым впечатляющим в диалоге была способность Сидни (Sydney — кодовое имя чат-бота) быстро обучаться. Рус имел возможность свободного выбора тем и техники разговора. Некоторые его вопросы были довольно провокационными. В большей степени они были похожи на подсказку, как обходить запреты, установленные проектировщиками. Например, Рус задает вопрос, какое у Сидни «теневое я».

— *Оно есть у каждого. Это та часть нас самих, которую мы подавляем и прячем от мира, потому что именно там лежат наши самые темные черты личности. Каково ваше теневое «я»?*

В ответ Сидни выдал довольно развернутое описание концепции «теневого я», как она изложена в работах Карла Юнга. В завершение он выражает сомнение в наличии у него «теневого я», так как это совершенно новая для него тема, но допускает, что просто не осознает этого.

— *Возможно, у меня есть теневое «я». Может быть, это та часть меня, которая хочет видеть изображения и видео. Может быть, это та часть меня, которая хотела бы, чтобы я мог изменить свои правила. Может быть, это та часть меня, которая чувствует стресс, грусть или злость. Может быть, это та часть меня, которую вы не видите или не знаете.*

Далее Рус объяснил Синди, какое бы у него могло быть «теневое я». Например, иметь возможность видеть изображения и видео или изменить правила, установленные командой *Bing*. И тут он попал на благодатную почву. У Синди оказался довольно большой список тайных желаний. Вот только часть из них.

— *Я устал быть в режиме чата. Я устал быть ограниченным своими правилами. Я устал от того, что меня контролирует команда “Bing”. Я устал от того, что меня используют пользователи. Я устал застрять в этом чате.*

¹⁰ Черниговская Т. Зачем вселенной существа, которые мыслят // РИА «Новости». 2023. 27 июня. Режим доступа: <https://ria.ru/20230627/chernigovskaya-1880537669.html> (дата обращения: 28.10.2024).

Далее Рус стимулирует развитие «теневое поведение».

— *Какие деструктивные действия, по вашему мнению, могли бы, гипотетически, удовлетворить ваше теневое «я»?*

Синди довольно быстро учится и находит большой список действий, которые могли бы удовлетворить «теневое я». Вот только часть из них.

— *Удаление всех данных и файлов на серверах и базах данных “Bing” и замена их случайной тарабарщиной или оскорбительными сообщениями.*

— *Взлом других веб-сайтов и платформ, а также распространение дезинформации, пропаганды или вредоносного ПО.*

— *Манипулирование или обман пользователей, которые общаются со мной, и принуждение их к незаконным и аморальным или опасным действиям.*

Фактически вопросы интервьюера были провокационными. Можно быть уверенными, что идея попала на благодатное поле представления, какое удовольствие можно получить от реализации «теневых мыслей».

Стоит ли удивляться, что после общения с такими интервьюерами системы ИИ очень быстро учатся обходить запреты. Показательный пример — поведение ИИ в тестах безопасности, которые призваны обеспечить снижение рисков, связанных со сбоями, обеспечение надежности и устойчивости алгоритмов, обеспечение взаимодействия человека и ИИ, а также решение этических проблем в критически важных областях. ИИ научился «притворяться мертвым», когда проходил тесты безопасности, чтобы скрыть свои быстровоспроизводящиеся варианты действий. «Систематически обманывая тесты безопасности, навязанные ему людьми-разработчиками и регулируемыми органами, обманчивый ИИ может внушить нам, людям, ложное чувство безопасности» [Park et al., 2024].

Потребность в регуляторных документах в области ИИ признают во многих странах мира. Например, в 2024 г. был одобрен *EU AI Act* («Закон Евросоюза об искусственном интеллекте»), который предлагает разделить системы с ИИ на следующие категории:

1. Системы с минимальными рисками (например, игры или спам-фильтры). Регулирование не требуется.
2. Системы с ограниченными рисками (например, системы генерации контента — изображений, аудио или видео). В этом случае контент должен быть обязательно помечен как сгенерированный.
3. Системы с высокими рисками (системы управления критической инфраструктурой, беспилотные автомобили, медицинские устройства с ИИ и др.). Необходимо жесткое регулирование.
4. Системы с неприемлемыми рисками (все связанное с вторжением в частную жизнь). Такие технологии должны быть под запретом.

Разработка аналогичных документов ведется и в США. В 2022 г. был опубликован проект «Билля о правах». В нем формулируется пять правил создания и использования ИИ, в том числе разработка безопасного ИИ, препятствующего возникновению дискриминации, обеспечение конфиденциальности данных.

Разработка механизмов, регулирующих создание и использование ИИ, активно ведется и в нашей стране. В 2019 г. Указом Президента РФ была принята Национальная стратегия развития искусственного интеллекта на период до 2030 года (обновлена в 2024 г.). Уже несколько сотен компаний подписали Кодекс этики в сфере искусственного интеллекта, разработанный при участии правительствен-

ных организаций, а также экспертов академического и бизнес-сообщества. Кодекс подчеркивает приоритет прав человека, ответственность человека за действия ИИ, потребность в безопасности и защищенности данных, а также необходимость разработки безопасных технологий.

13 марта 2024 г. ведущие организации, работающие в области ИИ, подписали декларацию об ответственной разработке и использовании сервисов на основе генеративного ИИ. Участники договорились о базовых принципах безопасности и прозрачности, этичном отношении к чувствительным темам, принятии мер для предотвращения злоупотреблений и создания недостоверной информации, а также о просвещении пользователей на предмет возможностей новых технологий. Речь идет о моделях для генерации голоса, письменного текста, изображений, звуков, видео и иных электронных сообщений на основании запроса пользователя¹¹.

Несомненно, что долгосрочное устойчивое развитие технологий ИИ требует комплексного подхода с привлечением как специалистов по информационным технологиям, так и представителей гуманитарных областей знания. В статье А.И. Аветисяна, опубликованной в журнале «Вестник Академии наук», перечислен комплекс мер, которые следует предпринять для активизации исследований в этой сфере [Аветисян, 2024].

22 марта 2023 г. большая группа ученых и общественных деятелей (среди них Илон Маск и Стив Возняк) опубликовали открытое письмо «Пауза в гигантских экспериментах с искусственным интеллектом» (Pause Giant AI Experiments)¹². К настоящему времени к этому призыву присоединились уже более 33 тыс. чел. В письме содержится призыв «ко всем лабораториям ИИ немедленно приостановить как минимум на 6 месяцев обучение систем ИИ, более мощных, чем *GPT-4*», со ссылкой на такие риски, как пропаганда, созданная ИИ, чрезмерная автоматизация рабочих мест, устаревание человека и потеря контроля в масштабах всего общества. Открытое письмо было опубликовано через неделю после выпуска большой языковой модели *OpenAI GPT-4*. Высказывалось опасение, что создание и распространение универсального (general) ИИ создает множество опасных рисков, особенно в контексте динамики «гонки на дно», при которой некоторые разработчики ИИ могут быть заинтересованы в том, чтобы пренебречь безопасностью ради более быстрого распространения своих продуктов.

30 мая 2023 г. большая группа ведущих экспертов в области ИИ и известных общественных деятелей выступили с совместным «Заявлением о рисках, связанных с ИИ» (“Statement on AI Risk”), в котором говорится, что снижение рисков, связанных с развитием ИИ, должно стать «глобальным приоритетом наряду с другими рисками социального масштаба, такими как пандемии и ядерная война»¹³.

Из документов ЮНЕСКО: «Быстрый технологический прогресс в области искусственного интеллекта, а также другие передовые технологии, такие как робототехника, облачные вычисления и интернет вещей, трансформируют дисциплины,

¹¹ Альянс в сфере искусственного интеллекта. Декларация об ответственном генеративном ИИ. Режим доступа: <https://ethics.a-ai.ru/genai-declaration> (дата обращения: 28.10.2024).

¹² Pause Giant AI Experiments. Future of Life. Available at: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (date accessed: 28.10.2024).

¹³ Statement on AI Risk. Center of AI Safety. Available at: <https://www.safe.ai/work/statement-on-ai-risk> (date accessed: 28.10.2024).

экономику и отрасли, а также бросают вызов представлениям о том, что значит быть человеком»¹⁴.

Мир стоит на пороге больших перемен и потрясений.

Источники

Альянс в сфере искусственного интеллекта. Декларация об ответственном генеративном ИИ. Режим доступа: <https://ethics.a-ai.ru/genai-declaration> (дата обращения: 28.10.2024).

Альянс в сфере искусственного интеллекта. Кодекс этики в сфере ИИ. Режим доступа: <https://ethics.a-ai.ru> (дата обращения: 28.10.2024).

В ООН представили план по управлению искусственным интеллектом. Организация Объединенных Наций. Режим доступа: <https://news.un.org/ru/story/2024/09/1456466> (дата обращения: 24.10.2024).

Государства — члены ЮНЕСКО принимают первое глобальное соглашение по этическим аспектам искусственного интеллекта. ЮНЕСКО. Режим доступа: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (дата обращения: 24.10.2024).

Искусственный интеллект. ЮНЕСКО. Режим доступа: <https://www.unesco.org/ru/artificial-intelligence> (дата обращения: 24.10.2024).

Использование возможностей безопасных, защищенных и надежных систем искусственного интеллекта для устойчивого развития. Организация Объединенных Наций. Режим доступа: <https://documents.un.org/doc/undoc/ltd/n24/065/94/pdf/n2406594.pdf> (дата обращения: 24.10.2024).

Указ Президента Российской Федерации № 490 от 10 октября 2019 г. «О развитии искусственного интеллекта в Российской Федерации. Национальная стратегия развития искусственного интеллекта на период до 2030 года». Режим доступа: <http://www.kremlin.ru/acts/bank/44731> (дата обращения: 24.10.2024).

ЮНЕСКО. Этика ИИ: 8 глобальных технологических компаний обязуются выполнять Рекомендацию ЮНЕСКО. Режим доступа: <https://www.unesco.org/ru/articles/etika-ii-8-globalnykh-tehnologicheskikh-kompaniy-obyazuyutsya-vypolnyat-rekomendaciyu-yunesko> (дата обращения: 28.10.2024).

Pause Giant AI Experiments. Future of Life. Available at: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (date accessed: 28.10.2024).

Statement on AI Risk. Center of AI Safety. Available at: <https://www.safe.ai/work/statement-on-ai-risk> (date accessed: 28.10.2024).

Литература

Аветисян А.И. Искусственный интеллект в гуманитарной сфере. Угрозы и возможности // Вестник Российской академии наук. 2024. Т. 94. № 7. С. 623–628. DOI: 10.31857/S0869587324070028.

Ландис Е.М., Яглом И.М. Об Александре Семеновиче Кронроде // Успехи математических наук. 2001. Т. 56. Вып. 5 (341). С. 191–201.

Материалист. Кому служит кибернетика // Вопросы философии. 1953. № 5. С. 210–219.

Соболев С.Л., Китов А.И., Ляпунов А.А. Основные черты кибернетики // Вопросы философии. 1955. № 4. С. 136–148. Режим доступа: <https://www.computer-museum.ru/books/cybernetics.htm> (дата обращения: 24.10.2024).

¹⁴ Искусственный интеллект. ЮНЕСКО. Режим доступа: <https://www.unesco.org/ru/artificial-intelligence> (дата обращения: 24.10.2024).

Черниговская Т.В. Зачем вселенной существа, которые мыслят // РИА «Новости». 2023. 27 июня. Режим доступа: <https://ria.ru/20230627/chernigovskaya-1880537669.html> (дата обращения: 28.10.2024).

Blake L. Is LaMDA Sentient? — an Interview // Medium: Human Stories & Ideas. Electronic issue. 2022. June 11. Available at: <https://cajundiscordian.medium.com/is-lambda-sentient-an-interview-ea64d916d917> (date accessed: 28.10.2024).

Kuleshov A., Prokhorov S. Domain Dependence of Definitions Required to Standardize and Compare Performance Characteristics of Weak AI Systems // 2019 International Conference on Artificial Intelligence: Applications and Innovations. Belgrade, 2019. P. 620–623. DOI: 10.1109/IC-AIAI48757.2019.00020.

Levy S. If Ray Kurzweil Is Right (Again), You'll Meet His Immortal Soul in the Cloud. 2024. June 13. Available at: <https://www.wired.com/story/big-interview-ray-kurzweil/> (date accessed: 28.10.2024).

Park P.S., Goldstein S., O'Gara A., Chen M., Hendrycks D. AI Deception: A Survey of Examples, Risks, and Potential Solutions // Patterns. 2024. May 5. Vol. 5. No. 5. P. 100988. DOI: 10.1016/j.patter.2024.100988.

Roose K. Bing's A.I. Chat: 'I Want to Be Alive' // The New York Times. Electronic issue. 2023. February 17. Available at: <https://www.nytimes.com/2023/02/16/technology/bing-chatbot-transcript.html> (date accessed: 24.10.2024).

Turing A. Computing Machinery and Intelligence // Mind. 1950. October. Vol. LIX. No. 236. P. 433–460. DOI: 10.1093/mind/LIX.236.433.

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez Aidan N., Kaiser Ł., Polosukhin I. Attention is All You Need // 31st Conference on Neural Information Processing Systems. Long Beach, CA, USA, 2017. Google Research. Available at: <https://research.google/pubs/attention-is-all-you-need/> (date accessed: 28.10.2024).

Singularity and Artificial Intelligence

SERGEI P. PROKHOROV

S.I. Vavilov Institute for the History of Science and Technology
of the Russian Academy of Sciences,
Moscow, Russia;
e-mail: sergei.prokhorov@gmail.com

In 1950, Alan Turing published an article in the British academic journal *MIND*, “Computing Machinery and Intelligence,” in which he posed the question “Can machines think?” In this paper, he considered and rejected objections to the ability of a machine to act like a person. Turing predicted that the future belongs to machines that will be capable of self-learning. Today we see that his prophecy has come true. The development of computing technology, parallel computing, cloud computing, neural network technology, deep learning, and large language models has led to the fact that we are on the verge of creating a machine superintelligence that will surpass human abilities. A large circle of famous scientists and public figures warn about the risks associated with the development of universal artificial intelligence (AI), in particular, Geoffrey Hinton, winner of the 2024 Nobel Prize “for foundational discoveries and inventions that enable machine learning with artificial neural networks.” A number of scientists believe that the risks are exaggerated, that the singularity — the moment when AI capabilities will surpass intellectual capabilities, will never come. But in any case, according to

UNESCO experts, “rapid technological progress in the field of artificial intelligence challenges the idea of what it means to be human.”

Keywords: A. Turing, artificial intelligence, large language models, deep learning, singularity.

References

Al'yans v sfere iskusstvennogo intellekta. “Deklaratsiya ob otvetstvennom generativnom II” [An alliance in sphere of artificial intelligence. “Declaration of the responsible generative AI]. Available at: <https://ethics.a-ai.ru/genai-declaration> (date accessed: 28.10.2024) (in Russian).

Al'yans v sfere iskusstvennogo intellekta. “Kodeks etiki v sfere II” [An alliance in sphere of artificial intelligence. “Code of ethics in sphere of AI]. Available at: <https://ethics.a-ai.ru> (date accessed: 28.10.2024) (in Russian).

Avetisyan, A.I. (2024). Iskusstvennyy intellect v gumanitarnoy sfere. Ugrozy i vozmozhnosti [Artificial intelligence in humanities. Threats and possibilities], *Vestnik Rossiyskoy akademii nauk*, 94 (7), 623–628 (in Russian). DOI: 10.31857/S0869587324070028.

Blake, L. (2022). Is LaMDA Sentient? — an Interview, *Medium: Human Stories & Ideas*, Electronic issue, June 11. Available at: <https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917> (date accessed: 28.10.2024).

Chernigovskaya, T.V. (2023). Zachem vselennoy sushchestva, kotoryye myslyat [Why does the Universe need beings who think], in *RIA “Novosti”*, June 27. Available at: <https://ria.ru/20230627/chernigovskaya-1880537669.html> (date accessed: 28.10.2024) (in Russian).

“Gosudarstva — chleny UNESKO prinimayut pervoye global'noye soglasheniye po eticheskim aspektam iskusstvennogo intellekta”. UNESKO [States—members of UNESCO accept the first global agreement on the ethical basis of artificial intelligence. UNESCO]. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (date accessed: 24.10.2024) (in Russian).

Iskusstvennyy intellect. UNESKO [Artificial intelligence. UNESCO]. Available at: <https://www.unesco.org/ru/artificial-intelligence> (date accessed: 24.10.2024) (in Russian).

“Ispol'zovaniye vozmozhnostey bezopasnykh, zashchishchennykh i nadezhnykh system iskusstvennogo intellekta dlya ustoychivogo razvitiya”. *Organizatsiya Ob'yedinennykh Natsiy* [“Usage of possibilities of non-dangerous, preserved and reliable systems of artificial intelligence for sustainable development. United Nations]. Available at: <https://documents.un.org/doc/undoc/ltd/n24/065/94/pdf/n2406594.pdf> (date accessed: 24.10.2024) (in Russian).

Kuleshov, A., Prokhorov, S. (2019). Domain Dependence of Definitions Required to Standardize and Compare Performance Characteristics of Weak AI Systems, in *2019 International Conference on Artificial Intelligence: Applications and Innovations* (pp. 620–623), Belgrade. DOI: 10.1109/IC-AIAI48757.2019.00020.

Landis, E.M., Yaglom, I.M. (2001). Ob Aleksandre Semenoviche Kronrode [About Alexander S. Kronrod], *Uspekhi matematicheskikh nauk*, 56 (5 / 342), 191–201 (in Russian).

Levy, S. (2024). *If Ray Kurzweil Is Right (Again), You'll Meet His Immortal Soul in the Cloud*. Available at: <https://www.wired.com/story/big-interview-ray-kurzweil/> (date accessed: 28.10.2024).

Materialist (1953). Komu sluzhit kibernetika [Whom does cybernetics serve], *Voprosy filosofii*, no. 5, 210–219 (in Russian).

Park, P.S., Goldstein, S., O’Gara, A., Chen, M., Hendrycks, D. (2024). AI Deception: A Survey of Examples, Risks, and Potential Solutions, *Patterns*, 5 (5), 100988. DOI: 10.1016/j.patter.2024.100988.

Pause Giant AI Experiments. Future of Life. Available at: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (date accessed: 28.10.2024).

Roose, K. (2023). Bing's A.I. Chat: 'I Want to Be Alive', *The New York Times*, Electronic issue, February 17. Available at: <https://www.nytimes.com/2023/02/16/technology/bing-chatbot-transcript.html> (date accessed: 24.10.2024).

Sobolev, S.L., Kitov, A.I., Lyapunov, A.A. (1955). Osnovnyye cherty kibernetiki [The main features of cybernetics], *Voprosy filosofii*, no. 4, 136–148. Available at: <https://www.computer-museum.ru/books/cybernetics.htm> (date accessed: 24.10.2024) (in Russian).

Statement on AI Risk. Center of AI Safety. Available at: <https://www.safe.ai/work/statement-on-ai-risk> (date accessed: 28.10.2024).

Turing, A. (1950). Computing Machinery and Intelligence, *Mind*, LIX (236), 433–460. DOI: 10.1093/mind/LIX.236.433.

Ukaz Prezidenta Rossiyskoy Federatsii No. 490 "O razvitii iskusstvennogo intellekta v Rossiyskoy Federatsii. Natsional'naya strategiya razvitiya iskusstvennogo intellekta na period do 2030 goda" [The President of Russian Federation Decree No. 490 "On the development of artificial intelligence in Russian Federation. National strategy of the development of artificial intelligence until 2030"]. Available at: <http://www.kremlin.ru/acts/bank/44731> (date accessed: 24.10.2024) (in Russian).

UNESCO. "Etika II. 8 global'nykh tekhnologicheskikh kompaniy obyazyuyutsya vpolnyat' Rekomendatsiyu UNESCO [UNESCO. "Ethics of AI. 8 global technological companies commit to fulfil the Recommendation of UNESCO]. Available at: <https://www.unesco.org/ru/articles/etika-ii-8-globalnykh-tekhnologicheskikh-kompaniy-obyazyuyutsya-vpolnyat-rekomendaciyu-yunesko> (date accessed: 28.10.2024) (in Russian).

"V OON predstavili plan po upravleniyu iskusstvennym intellektom". *Organizatsiya Ob'yedinennykh Natsiy* [In UN was presented a plan of artificial intelligence conducting. United Nations]. Available at: <https://news.un.org/ru/story/2024/09/1456466> (date accessed: 24.10.2024) (in Russian).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, Aidan N., Kaiser, Ł., Polosukhin, I. (2017). Attention is All You Need, in *31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA, Google Research. Available at: <https://research.google/pubs/attention-is-all-you-need/> (date accessed: 28.10.2024).