

**Искусственный интеллект для науки
и наука для искусственного интеллекта***

© 2022 г. К.В. Анохин^{1**}, К.С. Новоселов^{2***}, С.К. Смирнов^{3****},
А.Р. Ефимов^{4*****}, Ф.М. Матвеев^{5*****}

¹ Институт перспективных исследований мозга,
Москва, 119192, Ломоносовский пр., д. 27, корп. 1.

² Институт интеллектуальных функциональных материалов,
Национальный университет Сингапура, 117544, Singapore.

³ Женевский университет, Женева 4, 1211, ул. Женераль Дюфур, д. 24; Санкт-Петербургский
государственный университет, Исследовательская лаборатория им. П. Л. Чебышева,
Санкт-Петербург, 199178, 14 линия В.О., д. 29Б; Сколковский институт науки и технологий,
Центр перспективных исследований, Москва, 121205, Большой бульвар, д. 30/1.

⁴ ПАО Сбербанк, Москва, 121165, Кутузовский проспект, д. 32Б.

⁵ Московский государственный университет им. М.В. Ломоносова,
Москва, 119991, Ломоносовский проспект, д. 27, корп. 4.

** E-mail: k.anokhin@gmail.com

*** E-mail: kostya@nus.edu.sg

**** E-mail: sksmirnov@gmail.com

***** E-mail: makkawity@gmail.com

***** E-mail: matveevphil@gmail.com

Поступила 06.12.2021

В 2016 г. Хироаки Китано предположил, что искусственный интеллект (ИИ) сможет преодолеть ряд человеческих когнитивных ограничений, мешающих ускорить процесс достижения научных открытий [Kitano 2016 web]. С этого времени вопрос о возможности получения ИИ Нобелевской премии широко обсуждается в том числе и в академических кругах [Engineering for Research Symposium web 2020]. На конференции *AI Journey 2021* этот вопрос обсуждался представителями четырех научных дисциплин (физика, математика, нейробиология, философия), ставших соавторами данной статьи [AI 2021 web]. В этой научной дискуссии и статье, которая написана по ее следам, мы подвергли критическому анализу роль технологий ИИ в естественнонаучных исследованиях: насколько они могут быть полезны фундаментальной науке, каков потенциал или, наоборот, принципиальные ограничения использования ИИ в естественных и точных науках. Также часть нашей статьи посвящена встречному вопросу о том, что наука может дать будущим исследованиям ИИ. Сейчас невозможно представить себе машинное обучение без линейной алгебры, физики материалов, исследований

* Ф. Матвеев выполнил работу в рамках финансирования Российского научного фонда. Проект 21-18-00184 «Социально-гуманитарные основы критерии оценки инноваций с использованием цифровых технологий и искусственный интеллект».

К. Новоселов выполнил работу в рамках финансирования Министерством образования Сингапура программы Центра исследований компетенций. Проект EDUNC-33-18-279-V12.

мозга – все это лежит в основе того, что сегодня принято объединять под общим «зонтичным» термином ИИ [Russel, Norvig 2021]. Таким образом, раз послужив рождению ИИ, чем физика, математика и нейробиология могут послужить ему сегодня?

Ключевые слова: искусственный интеллект, нейробиология, математика, физика.

DOI: 10.21146/0042-8744-2022-3-93-105

Цитирование: Анохин К.В., Новоселов К.С., Смирнов С.К., Ефимов А.Р., Матвеев Ф.М. Искусственный интеллект для науки и наука для искусственного интеллекта // Вопросы философии. 2022. № 3. С. 93–105.

AI for Science and Science for AI*

© 2022 Konstantin V. Anokhin^{1**}, Konstantin S. Novoselov^{2***}, Stanislav K. Smirnov^{3****},
Albert R. Efimov^{4*****}, Philipp M. Matveev^{5*****}

¹ *Institute for Advanced Brain Studies,
27/1, Lomonosovsky av., Moscow, 119192, Russian Federation.*

² *Institute for Functional Intelligent Materials, National University of Singapore, 117544, Singapore.*

³ *University of Geneva, 24, rue du Général Dufour, 1211, Geneva 4, Switzerland; Saint Petersburg State University, Chebyshev Laboratory, 29B, 14 line of the VO, Saint Petersburg, 199178, Russian Federation; Skolkovo Institute of Science and Technology, Center for Advanced Studies, 30/1, Bolshoy Boulevard, Moscow, 121205, Russian Federation.*

⁴ *OJSV Sberbank, 32B, Koutouzovsky av., Moscow, 121165, Russian Federation.*

⁵ *Lomonosov Moscow State University,
27/4, Lomonosovsky av., Moscow, 119991, Russian Federation.*

^{**} *E-mail: k.anokhin@gmail.com*

^{***} *E-mail: kostya@nus.edu.sg*

^{****} *E-mail: sksmirnov@gmail.com*

^{*****} *E-mail: makkawity@gmail.com*

^{*****} *E-mail: matveevphil@gmail.com*

Received 06.12.2021

In 2016, Hiroaki Kitano proposed that artificial intelligence (AI) will be able to overcome a number of human cognitive limitations that slow down the process of scientific discovery [Kitano 2016 web]. Since then, the odds of AI being awarded the Nobel Prize have been widely discussed, particularly within academic community [Engineering for Research Symposium web 2020]. At the *AI Journey 2021* conference, some renowned representatives of four scientific disciplines (physics, mathematics, neurobiology, philosophy) discussed this issue and then co-authored this article [AI 2021 web]. In the first part of our paper, we critically analyze the role of AI technologies in natural science research: how useful they can be for fundamental science, what the potential of AI in natural and exact sciences is, and what principal limitations it has. Another part of our article discusses a counter-question of what science can do for the future research into AI. Today, it is impossible to imagine machine learning without linear algebra, physics of materials, and brain research. All this falls under what is now commonly referred to as AI, a general umbrella term [Russel, Norvig 2021]. Thus, having served the birth of AI once, how can physics, mathematics, and neuroscience serve it today?

Keywords: artificial intelligence, neuroscience, mathematics, physics.

DOI: 10.21146/0042-8744-2022-3-93-105

* P. Matveev performed this work under the funding of the Russian Science Foundation. Project 21-18-00184 "Socio-humanitarian foundations criteria for evaluating innovations using digital technologies and artificial intelligence".

K. Novoselov performed this work under the funding by the Ministry of Education of Singapore through the Research Center of Excellence program. Project No. EDUNC-33-18-279- V12, I-FIM.

ИИ для математики

Галилео Галилей, подчеркивая связь математики с миром, говорил, что книга природы написана на языке математики. Неудивительно, что наше рассуждение мы начнем с использования ИИ в математике. Еще в 70-х гг. лауреат премии Филдса Пол Коэн заявлял, что «компьютеры рано или поздно заменят математиков» [Sriraman 2017]. В 2018 г. другой лауреат Филдсовской премии и один из авторов (Ст. Смирнов) заявлял о появлении различных исследовательских групп, которые занимаются исследованиями в этом направлении, отмечая прогресс: создана система, которая нашла и доказала половину теорем из первых трех книг Евклида [Beeson, Narboux, Wiedijk 2019]. Другим примером является достижение автоматического вывода доказательства в одном из формальных верификаторов (COQ) [Kaliszyk, Mamane, Urban 2014]. Кевин Базард запустил проект использования компьютерных систем проверки доказательств в чистой математике (*virtualstudent*). Питер Шольце разработал свой амбициозный план компьютеризационистской унификации математики [Castelvecchi 2021]. Группа *Google Research* под руководством Кристиана Сегеди попыталась использовать обработку естественного языка, чтобы сделать компьютерные доказательства более «человечными» по структуре и объяснению [Ornes 2021 web]. Как математики используют ИИ сейчас?

Авторы статьи видят два основных направления применения ИИ в области математики. Во-первых, это возможность обработки огромного набора научной информации, в котором ИИ помогает ориентироваться. Появление элементарного компьютерного доступа к оцифрованным базам значительно упростило работу математиков. Во-вторых, и здесь математика отличается от физики – это проверка результатов. Когда физик разрабатывает теорию, он проверяет ее с помощью лабораторных экспериментов, и если предсказания теории не выполняются, то она отбрасывается. Когда математик разрабатывает теорию, ему нужно лишь проверять доказательства – это мы усваиваем со школьной скамьи. Так, вышеупомянутые *Elementa* Евклида математики пытались проверять еще до появления статистического ИИ с помощью старых логических методов в стиле машины Тьюринга. Сейчас в этой области есть серьезный прогресс. Например, одну очень сложную теорему из теории групп, доказательство которой занимает более двухсот страниц, проверили с помощью логической системы [Gonthier, Asperti, Avigad, Bertot, Cohen, Garillot et al. 2013]. Прогресс в этом направлении серьезно обсуждается в ведущих математических журналах [Buzzard 2020] и на Международных конгрессах математиков [Davenport, Poonen, Maynard, Helfgott, Tiep, Cruz-Filipe 2018 web]. Теоретически почти любое доказательство сейчас может быть проверено таким образом. Единственная проблема, которая возникает в данном контексте, – это необходимость выражения доказательства на формальном языке, схожем с современными языками программирования, которое необходимо во избежание двусмысленностей естественного языка. Поэтому возникает сложность перевода с естественного языка на универсальный язык формальной проверки доказательств: на данный момент математики не умеют делать этого достаточно хорошо. И здесь мы выходим на более сложную проблематику *Natural Language Processing*: человек-математик не всегда способен осуществить перевод на формальный язык работы другого математика.

Помимо двух вышеозначенных направлений, говоря о перспективах ИИ в математике, один из авторов (Ст. Смирнов) выделяет и более интересную проблему – может ли ИИ самостоятельно искать новые теоремы. Математики часто сетуют на то, что математика стала слишком сложна: через пару десятков лет доказательства теорем достигнут такой длины, с которой человеческий ум будет не в состоянии справляться. Но несмотря на почти столетние рассуждения такого рода, большой прогресс во многих

областях математики до сих пор есть: за последние 25 лет в математике было совершенно много открытий, включая доказательство гипотезы Пуанкаре и теоремы Ферма. Если человеческий предел в математике когда-то все-таки будет достигнут, то станет актуальным вопрос о том, сможет ли за него выйти ИИ.

Здесь речь может идти как о доказательствах известных теорем, так и о самом поиске новых теорем. Вспомним, что современная математика и физика начались с того, что Тихо Браге пронаблюдал закономерности движения планет, которые проанализировал Иоганн Кеплер, заметив, что планеты движутся не по окружностям, а по эллипсам с очень маленьким эксцентриситетом. Уже после того, как эти закономерности были описаны, Исаак Ньютон придумал, каким образом можно вывести их из простого уравнения. Если бы эти ученые имели доступ к компьютеру, то ИИ, скорее всего, гораздо быстрее нашел бы те закономерности, которые они вынуждены были пытаться обнаружить «на глаз». А теперь давайте представим, что существуют закономерности, которые настолько сложны, что человек в принципе не в состоянии их заметить: скажем, можно представить, что мы не можем объединить квантовую механику и теорию относительности из-за того, что в этих областях существует некая закономерность, которую мы просто не способны обнаружить. В этом случае кажется, что ИИ, который превзойдет нас в когнитивных способностях, может оказаться способным заметить такую важную закономерность. Другой вопрос, сможем ли мы, в свою очередь, понять, что ИИ заметил эту закономерность. Пока что у нас нет ответа на эти вопросы, но академическое сообщество ожидает прогресса в этом направлении в ближайшие годы, и этому есть все основания. Существуют уже несколько интересных примеров того, как ИИ помог найти нетривиальные закономерности [Davies, Veličković, Buesing, Blackwell, Zheng, Tomašev, Tanburn, Battaglia, Blundell, Juhász, Lackenby, Williamson, Hassabis, Kohli 2021], но пока все они сводятся к случаям, где нет принципиальных сложностей, помимо слишком большого для человека объема информации. Некоторые из этих результатов будут обсуждаться на следующем Международном конгрессе математиков.

Необходимо оговориться, что в данном случае наши рассуждения только гипотетические: уровень математической сложности, который человек не может преодолеть, пока еще не достигнут, и, более того, мы не можем с уверенностью сказать, что он вообще существует. На данный момент математики успешно выдвигают новые гипотезы и теоремы, поэтому возможно, что этот предел скорее следует искать в физике и в других естественных науках, где он может быть более наглядным. В качестве иллюстрации предположим, что в одном из возможных миров притяжение двух тел зависело бы от четности количества атомов в этих телах: кажется, обладая достаточным количеством необходимых данных, ИИ смог бы заметить это с гораздо большей вероятностью, чем человек.

ИИ для физики

Многие методы ИИ используются в теоретической физике. К примеру, машинное обучение применяется в физике элементарных частиц [Leeming 2021]. В 2021 г. был составлен сборник методов ИИ в физике элементарных частиц, включающий более 350 статей [Feickert, Nachman 2021]. Один из авторов (К. Новоселов) отмечает, что нейросеть влияет на разработку материалов «самым прямым образом», позволяя «планировать эксперименты и условия синтеза новых материалов, исходя из имеющихся данных» (см. об этом: [Скрынников 2021 web]). «И в этом отношении она становится в некотором смысле конкурентом человеку» [Там же]. Закономерно возникает вопрос о дальнейшем развитии применения ИИ в физике: ограничится ли «конкуренция» с человеком внедрением вычислительных методов в экспериментальную область с целью повышения производительности или возможны некие альтернативные способы применения ИИ.

Авторы полагают, что включение методов ИИ в фундаментальную физику уже произошло: ИИ помогает решать практические и иногда даже фундаментальные вопросы. Это объясняется просто: физики в целом не любят громоздких и сложных задач, то есть задач, которые не описываются небольшим количеством красивых уравнений. Поэтому,

если возникает необходимость написать уравнение в полстраницы, физик скорее всего решит, что в его рассуждении что-то пошло не так. Отсюда возникает достаточно большой класс задач, которыми пренебрегают физики-люди. И это именно та область, где ИИ может помочь нам найти некие закономерности в работе с «некрасивыми» громоздкими уравнениями, ведь помимо красоты основной задачей физики все-таки остается получение знания о фундаментальных естественных процессах и предсказание на основе этого знания поведения физических структур. Таким образом, ИИ может помочь и в предсказании, и в символическом описании физических процессов.

С точки зрения одного из авторов (К. Новоселова), одна из наиболее интересных областей, где ИИ может существенным и нетривиальным способом помочь физике – создание языка теории. Как уже было сказано, иногда физики получают «некрасивые» уравнения, и в этом случае они склонны решать возникшую проблему двумя способами: либо просто отбрасывая задачу, либо придумывая новый язык описания, редуцирующий задачу. Например, ученым было очень сложно догадаться, что в уравнении Шредингера необходимо ввести волновую функцию: само это введение в начале XX в. было тяжелым, трудоемким и проблемным процессом. Но как только она была введена, появились достаточно простые мнемонические правила, которые сегодня осваиваются любым студентом первого курса. Таким образом, введение нового описательного языка – это очень полезная, хотя и трудоемкая процедура.

Роль ИИ заключается в конструировании нового языка, описывающего в простых терминах и уравнениях смысл замеченных им закономерностей. По мнению одного из авторов (К. Новоселов), наука подошла к тому моменту, когда ИИ способен это осуществить. Например, в недавнем исследовании *Deep Mind* о свертывании белка, входные графы модифицировались в зависимости от выходных данных, что было правильным направлением создания алгоритма так, чтобы получить не только взаимосвязь между входными и выходными параметрами, но и адекватное физическое описание для такой системы.

ИИ для нейробиологии

Переходя от физики к нейробиологии, заметим, что нейробиологическое исследование обработки информации в человеческом мозге фактически вдохновило развитие машинного обучения. Каким образом ИИ дает нейробиологам возможности получать новые знания о механизмах работы нашего мозга? В этой связи интересно вернуться к статье доктора Китано, который отмечает особую полезность внедрения ИИ именно для биомедицинских наук [Kitano 2016 web]. Нейроученые, занимающиеся вычислительной нейробиологией, часто заявляют, что ИИ можно использовать не только в качестве инструмента для обработки сложных наборов данных, но и в качестве модели для разработки и тестирования идей о том, как наш мозг выполняет вычисления [Savage 2019]. Соответственно, возникает следующий вопрос: существует ли в нейробиологии специфическое поле применения ИИ, который будет не только оптимизировать работу с большими данными, но и помогать делать новые открытия, которые могут быть удостоены Нобелевской премии?

Прежде всего один из авторов (К. Анохин) хотел бы прокомментировать рассуждение о Нобелевской премии для ИИ: это представляется маловероятным. Если уровень развития ИИ позволит ему делать открытия уровня Нобелевской премии, то за эти открытия просто перестанут ее присуждать, фокус природы научного открытия будет смещен от каких-то более технологических вещей в сторону более креативных. Ранее Нобелевскую премию присуждали за расшифровку структуры белков [Sanger 1959], но по мере автоматизации такие достижения не будут больше ее предметом. Он сместится в более «темные» области, решение проблем в которых недостаточно алгоритмизировано. Современные вариации ИИ не попадают в один из трех важнейших методов ученого, выделенных Чарльзом Пирсом, – абдукцию, которую американский логик определял как процесс формирования объяснительных гипотез, единственную

логическую операцию, которая вводит новую идею [Peirce 1974, 106]. В работе ученые пользуются индукцией, дедукцией и абдукцией. Индукция – это сбор большого количества фактов и выведение из них неких закономерностей. Здесь хорошо применимы современные нейронные сети. Дедукция – это логический вывод, в котором хорошо применим логический ИИ. Но есть и третий метод, который Эйнштейн, много занимавшийся методологией науки, считал критическим моментом научного творчества и называл интуитивным охватом большого количества фактов или воображением, с помощью которого в истории и делались самые крупные научные открытия [Einstein, Shaw 2012, 97]. Он также считал, что решения просты, а самое трудное – это открыть проблему [Howard, Giovanelli 2019 web]. В 1980-е гг. в ИИ возникло направление, получившее название «абдуктивное логическое программирование» [Denecker, Kakas 2000]. Но оно опиралось скорее на методы дедукции, природа же истинного абдуктивного вывода остается пока неформализованной [Larson 2021]. Поэтому, отвечая на вопрос о Нобелевской премии для ИИ, мы полагаем, что большая область самых сложных проблем все-таки останется за человеком – это поиск в темноте, интуитивный и трудно формализуемый процесс выявления сути проблемы и алгоритма ее решения.

Сегодня получение многих глубоких результатов в области исследований мозга и интеллекта получает прирост благодаря методам машинного обучения, которые являются технологиями ИИ. Например, мы нашли ответ на вопрос о том, как именно мозг позволяет нам видеть окружающий мир и что именно «видят» нервные клетки нашего мозга. Понять это позволили методы ИИ и супернормальные стимулы. Супернормальным называется искусственный стимул, который вызывает у организма гораздо более сильную поведенческую реакцию, чем естественный. В классических работах Нико Тинбергена птенцу серебристой чайки показывали карандаш с красной меткой, которая визуально похожа, но крупнее, чем красное пятно на клюве матери, видя которое, птенец выпрашивает у нее пищу. Птенец реагировал на красный стержень гораздо сильнее, чем на натуральный клюв матери [Тинберген 1978]. Такой прием позволяет этологам узнавать, что же на самом деле наиболее важно организму в окружающем мире. Два года назад в лаборатории Маргарет Ливингстон в Гарварде исследователи разработали глубокую генеративную искусственную нейронную сеть под названием *XDREAM*, позволившую похожим способом узнать, что во внешнем мире наиболее важно уже не самому подопытному животному, а нейронам в его мозге. Сеть была спроектирована таким образом, чтобы синтезировать изображения, начиная с «чистого листа», и постепенно усложнять их. При этом каждый раз оценивалось, какотреагируют на это изображение нейроны в мозге обезьяны. Отбиралось изображение, вызывавшее самый сильный ответ, оно снова изменялось, снова отбиралось самое «привлекательное» для нейрона изображение, и так далее, много раз [Ponce, Xiao, Schade, Hartmann, Kreiman, Livingstone 2019]. С помощью этого приема максимизации ответа нейронов исследователям удалось выяснить, что на самом деле «видят» и «любят» нервные клетки в глубине мозга. В результате получились сложные синтетические изображения, в которых узнавались причудливо переплетенные отдельные признаки мира: например, изображение, склеенное из лица сотрудницы в маске, которая ухаживает за обезьяной в виварии, и кусок красного ошейника другой обезьяны, которая сидела в виварии в клетке напротив.

Таким образом, мы получаем окно в изучение глубоких когнитивных характеристик биологических нейронных сетей. Недавно это пионерское исследование получило неожиданное зеркальное продолжение в области свойств искусственных нейронных сетей. Известная компания *Open AI* применила этот же метод максимизации ответа для исследования того, что «видят» нейроны в глубоких слоях их новой нейросетевой структуры *CLIP* [Goh, Cammarata, Voss, Carter, Petrov, Schubert, Radford, Olah 2021 web]. Оказалось, что для рецептивных полей нейронов этой сети свойственна такая же странная картина склеенных из многих признаков, но узнаваемых изображений: некоторые нейроны имеют рецептивные поля на обобщенные признаки лица

человека, некоторые – на те или иные индивидуальные личности, например политических деятелей; причем происходит это так же абстрактно, как и в биологических клетках, то есть одна и та же клетка может «узнавать» как лицо, так и имя данного человека. Похожей технологией является *ru DALL-E*, которая генерирует изображение на основе введенного текста [Димитров 2021 web]. Итак, некоторые нейроны кодируют эмоциональное выражение, но как и почему искусственная нейронная сеть, которая во многом так непохожа на биологическую, способна формировать в своих глубоких слоях такие когнитивные специализации искусственных нейронов? В разрешении этого вопроса скрыт ключ к пониманию неких общих закономерностей кодирования и распознавания, стоящих за столь разными (естественными и искусственными) архитектурами.

Математика для ИИ

Современный ИИ возник не на пустом месте. Его истоки можно проследить в человеческой истории вплоть до античных времен [Vardi 2021 web]. Поэтому уместно сказать о том, какие знания могут дать точные и естественные науки для будущих прорывов в исследованиях ИИ. Многие математики (в том числе один из авторов, С. Смирнов) уверены, что в современных исследованиях ИИ практические результаты значительно обгоняют теоретическую базу: «Мы находимся в положении людей, которые строили паровозы до открытия законов термодинамики: паровозы ездили, но иногда взрывались. Результаты физики позволили сделать паровозы, которые ездили гораздо лучше и уже не взрывались» [Смирнов 2021 web]. Вместе с тем есть представление о том, что именно математическая детализация может дать серьезный прогресс области ИИ в ближайшие 10 лет, возможности появления некой новой науки об ИИ: «Возможно, это должна быть наука о структуре данных, возможно, это должен быть комплекс наук, кодирующих видеоряды, аудиосигналы, язык и т.д.» [Там же]. Соответственно, возникает вопрос о том, какую роль в этой новой науке об ИИ будет играть математика.

Возвращаясь к теме простого языка описания, приведем пример выдающегося исследователя Жана-Батиста Жозефа Фурье, который был ученым и высокопоставленным администратором наполеоновской империи. Изучая теплопроводность, Фурье проводил много экспериментов в своем кабинете: открывая и закрывая окно, он наблюдал за изменениями температуры. Затем он понял, что для того, чтобы решать уравнения распространения тепла, гораздо проще представить уравнения теплопроводности как суммы синусов, аналогично тому, как мы представляем звук суммой нот. Базис Фурье действительно позволяет гораздо лучше и проще представлять очень большое множество феноменов. Позднее описательный язык Фурье был улучшен: была придумана концепция вейвлетов, локальных аналогов синуса для решения задач в чистой математике. Первый базис непрерывных мини-волн построил Ян-Олав Стремберг. Уже на основе этого Ив Мейер и Рональд Койфман довели этот метод до совершенства применительно к кодированию данных. Результат мы используем сегодня как стандартный формат JPEG, в котором хранится большинство фотографий.

Итак, безусловно верная мысль о том, что нам нужен новый язык, зависит от того, чего именно мы хотим достичь: либо сделать что-то, что уже умеет делать человек, либо пойти куда-то дальше. В последнем случае мы сталкиваемся с эпистемологической проблемой «незнания о незнании» (*unknown unknowns*) или «допроблемной ситуацией» [Дубровский 2010, 167], когда непонятно, что, собственно, нужно искать и анализировать, ведь мы можем просто не знать о том, какие закономерности должны исследовать. Если же мы остаемся в пределах, доступных человеку, то возникает интересный момент в плане реализации: обычно мы пытаемся сделать что-то по аналогии с процессами в мире, но несколько по-другому [Schaeffer 2007 web]. В качестве примера из математики можно взять следующие задачи.

1. Допустим, Вам покажут экран телевизора в миллион пикселей, на котором будут беспорядочно перемешаны белые и черные пиксели, и спросят о том, каких пикселей больше. Человек с легкостью ответит на этот вопрос: в визуальном кортексе, уже на первом уровне, будут сформированы некоторые преобразования Фурье, которые помогут уверенно решить, каких пикселей больше. Если мы дадим эту задачу нейросети, она тоже довольно быстро научится ее решать.
2. Допустим, Вам покажут экран и спросят о том, четное или нечетное количество белых пикселей на нем изображено. В случае задачи «5 на 5» Вы посчитаете очень быстро, для «10 на 10» будет вероятность сбиться со счета, а в случае «1 млн на 1 млн» задача окажется непосильной. И, что интересно, нейросеть неспециализированной архитектуры точно так же не сможет решить эту задачу на основе разумного количества примеров.

Таким образом, первую задачу достаточно легко решает и человек, и нейросеть, а вторую задачу не могут решить оба. Первая задача маломерная, то есть определяется одной гиперплоскостью в миллионном пространстве параметров, а вторая в приведенной формулировке – очень многомерная. Видимо, то, что человек умеет делать, – это хорошо решать маломерные проблемы: мы смотрим на картинку в миллион пикселей и с легкостью сводим задачу в миллионмерном пространстве к структуре маломерной задачи. Примерно то же происходит и с нейросетями. Поэтому нам необходим язык, который описывает маломерные геометрические структуры многомерных пространств. Конечно, здесь нам мешает человеческая интуиция трехмерного пространства, ведь многомерные пространства ведут себя совсем по-другому. Это возвращает нас к построению «красивых» формул: мы комфортно чувствуем себя тогда, когда размерность того объекта, который мы изучаем, как минимум меньше 100, поэтому «формулу длиной в полстраницы» мы хотим редуцировать до чего-то более простого.

Интересно также, что в случае изобретения ИИ мы, гипотетически, сможем повысить размерность, которую он сможет эффективно оценивать: например, если человек может оценивать 100, то ИИ сможет оценить 200. В таком случае возникает резонная гипотеза о том, что он сможет придумать нечто новое, то, до чего мы не смогли бы додуматься.

Физика для ИИ

Как и нейробиология, физика часто ставила перед ИИ задачи, решение которых стимулировало развитие области: например, обнаружилось, что некоторые фундаментальные концепции из теоретической физики применимы в машинном обучении, как, например, было с машиной Больцмана, алгоритмом имитации отжига или степенным законом. Один из участников нашей дискуссии (К. Новоселов) подчеркивал важность новых интеллектуальных материалов с «нетривиальными свойствами» в создании нейроморфных компьютеров, которые работают по аналогии с мозгом, а также важность таких материалов для исследования интерфейсов с живыми объектами, в частности интерфейсов мозг-компьютер (см. об этом: [Бутрин 2021 web]). Кажется, что это открывает огромные перспективы для исследований мозга, но чем физика может помочь исследованиям ИИ сегодня?

Один из авторов (К. Новоселов) считает, что такая постановка вопроса затрагивает очень интересную проблему: сегодня ученые плохо понимают, как вообще должны работать нейроморфные компьютеры, ведь такие устройства будут нелинейны. Как только мы получаем нелинейную систему, мы перестаем ее понимать, так как не можем хорошо предсказывать результат. Но описываемые системы должны быть не только нелинейными, но и обладать памятью. А как только мы говорим про системы с памятью, мы автоматически подразумеваем, что это неравновесная система, то есть то, что физика даже не берется описывать: физики всегда пытаются описать простейшие равновесные системы и уже потом описывать поведение контролируемого вывода системы

из равновесия. Таким образом, нелинейная система с памятью – это огромный пласт задач, которые физике только предстоит решать. Тем не менее такие системы уже существуют, и физики примерно понимают, как следует с ними оперировать, чтобы получать эффекты памяти.

Как было отмечено выше, ИИ действительно применяется в физике для решения широкого круга задач, но попробуйте интегрировать ИИ в Ваш мобильный телефон и у Вас возникнут огромные проблемы: мы пытаемся решать существенно нелинейные нейроморфные задачи на обычных процессорах, которые для этого не предназначены. Поэтому, создавая материалы и устройства, которые обладают памятью, мы получаем возможность проводить вычисления уже на самих устройствах памяти, а не только на обычных микропроцессорах. Именно поэтому создание динамических материалов с памятью – это задача, которой сейчас занимаются многие физики. Но, с точки зрения авторов, проблема в том, что у нас нет способа описания подобных материалов: физика описывает материалы только в состоянии равновесия. И как раз для решения этой проблемы описания возможно использование динамических ИИ, которые «живут» во времени и которые могут помочь в описании динамических материалов для разработки нейроморфных компьютеров. Таким образом, ИИ в Лаборатории программируемых функциональных материалов пытается решить амбициозную задачу «расчета самой себя», создавая компьютеры для создания самой себя.

Нейробиология для ИИ

Завершающая часть обсуждения посвящена перспективам нейробиологии в применении к ИИ. Некоторые исследователи ИИ выражают мнение, что, будучи изначально вдохновлены нейробиологией, современные методы машинного обучения уже не ориентируются на биологические модели. Например, Джонатан Шеффер говорил, что, хотя подходы, вдохновленные биологией, полезны, они, вероятно, не будут иметь ничего общего с фактическими достижениями в исследованиях искусственного интеллекта, за исключением некоторых высокоуровневых мотиваций или абстракций [Schaeffer 2007 web]. В качестве иллюстрации своей мысли он приводил попытки человека воспроизвести полет. Хотя изучение птиц дает нам некоторые высокоуровневые идеи и фундаментальные концепции, самолеты, разработанные до сих пор, функционируют совсем не так, как птицы [Russell, Norvig 2021]. Другие исследователи считают, что нейробиология и ИИ неразрывно связаны [Savage 2019]. Один из участников дискуссии (К. Анохин) является руководителем ряда проектов, в которых нейробиологические исследования естественного интеллекта призваны помочь созданию искусственного. Полученный опыт и результаты этих исследований позволяют рассуждать о том, как современная нейробиология взаимодействует с актуальными исследованиями ИИ. Является ли она по-прежнему основным источником вдохновения и чем должен отличаться ИИ вдохновленный нейробиологией от других систем ИИ?

По мнению одного из авторов (К. Анохина), проекции нейронауки в ИИ скорее всего сохраняют свою значимость. Об этом свидетельствует хотя бы тот факт, что многие выдающиеся архитекторы ИИ были либо психологами, либо нейробиологами по образованию. И сегодня такие передовые компании в секторе ИИ, как, скажем, *Deep Mind*, в значительной степени строят свои разработки на нейронаучных исследованиях. Принципы, используемые в самолетостроении, не повторяют буквально то, что используют птицы, но есть огромный диапазон того, что умеет делать биологическая система и что пока не способна делать ни одна искусственная. Вряд ли кому-то приходилось слышать, чтобы воробей потерпел крушение при посадке. Поэтому область перспективных для ИИ нейробиологических возможностей очень велика. Например, память, о которой говорилось в связи с физикой, – это функция, которая широко используется в биологических нейронных сетях, но пока гораздо хуже работает в искусственных, которые требуют миллионов повторений для запоминания того, что человек запомнил бы с одного раза.

Один из участников дискуссии (К. Анохин) предполагает, что справедлива и обратная ситуация: искусственные системы способны заходить в области, в которых биологические системы пока еще очень несовершенны. В качестве примера можно взять способность находить причинные связи между далеко отстоящими друг от друга событиями. Установление причинных отношений между событиями в мире является одной из ключевых адаптивных функций интеллекта. Но биологический интеллект способен делать это лишь на очень ограниченном интервале времени. Это было открыто в Павловской школе и с тех пор многократно проверено: несколько десятков секунд для животного и несколько минут для человека – это интервал, за пределами которого интеллекту трудно установить отношения причинного следования между условным и безусловным сигналом [Павлов 1949]. Но искусственные нейронные сети потенциально способны устанавливать причинные отношения на гораздо более длинных интервалах: то есть они могут сделать то, что не могут делать биологические сети. Проблема в том, что им для этого требуются миллионы повторений, в то время как в биологических нейронных сетях выработка условных связей способна происходить с одного раза. Так что здесь как искусственному, так и естественному интеллекту есть чему поучиться друг у друга.

Заключение

ИИ обладает значительным потенциалом в научных исследованиях прежде всего за счет возможности анализировать скрытые от человека закономерности, предлагая дополнительные описательные возможности, от нового языка до новой науки, объединяющей различные дисциплины.

Фундаментальные точные и естественные науки, несомненно, играют огромную роль в передовых исследованиях ИИ, так как дают возможность лучше понять потенциал мозга человека и дают лучшие инструменты и способы реализации сложнейших интеллектуальных функций. Авторы считают, что наше понимание работы мозга находится даже не на уровне «черного ящика», так как на современном уровне знаний у науки нет понимания, что именно и на какой из входов «черного ящика» необходимо подать, чтобы получить результат.

Источники и переводы – *Primary Sources and Translations*

Павлов 1949 – Павлов И.П. Избранные произведения. М.: Изд-во Академии наук СССР, 1949 (Pavlov, Ivan P., *Selected works*, in Russian).

Тинберген 1978 – Тинберген Н. Поведение животных. М.: Мир, 1978 (Tinbergen, Nikolaas, *Animal Behaviour*, Russian Translation).

Peirce, Charles S. (1974) “Harvard Lectures on Pragmatism: Lecture VI”, Hartshorne, Charles, Weiss, Paul, eds., *Collected Papers of Charles Sanders Peirce*, Vol. V–VI, Harvard University Press, Cambridge, pp. 94–111.

Sanger, Frederick (1959) “Chemistry of insulin”, *Science*, Vol. 129 (3359), pp. 1340–1344.

Einstein, Albert, Shaw, Bernard (2012) *Einstein on cosmic religion and other opinions and aphorisms*, Courier Corporation, Chelmsford.

Ссылки – *References in Russian*

Димитров 2021 web – Димитров Д. «RuDALL-E: генерируем изображения по текстовому описанию, или Самый большой вычислительный проект в России». URL: <https://habr.com/ru/company/sberbank/blog/586926/>

Дубровский 2010 – Дубровский Д.И. Обман. Философско-психологический анализ. М.: Канон+: Реабилитация, 2010.

Скрынников 2021 web – Скрынников А. «Фантастически интересные свойства»: российский ученый о новых материалах и технологиях. URL: <https://ru.rt.com/hfg3>

Бутрин 2021 web – Бутрин Д. Константин Новоселов создает лабораторию «умных» материалов в МФТИ. URL: <https://www.kommersant.ru/doc/4683651>

Смирнов 2021 web – Смирнов С.К. Беседа с математиком. Публичное интервью со Станиславом Смирновым. URL: <https://youtu.be/lvFu6IMXwj0>

References

- AI web (2021) URL: <https://ai-journey.ru/en/media/broadcast#galery-video/773>
- Beeson, Michael, Narboux, Julien, Wiedijk, Freek (2019) “Proof-checking Euclid”, *Annals of Mathematics and Artificial Intelligence*, Vol. 85, No. 2–4, pp. 213–257.
- Butrin, Dmitry (2021) *Konstantin Novoselov creates a laboratory of “smart” materials at MIPT*, URL: <https://www.kommersant.ru/doc/4683651>
- Buzzard, Kevin (2020) “Proving Theorems with Computers”, *Notices of the American Mathematical Society*, Vol. 67, No. 11, pp. 1791–1799.
- Castelvecchi, Davide (2021) ‘Mathematicians welcome computer-assisted proof in “grand unification” theory’, *Nature*, Vol. 595, No. 7865, pp. 18–19.
- Davenport, James H., Poonen, Bjorn, Maynard, James, Helfgott, Harald, Tiep, Pham H., Cruz-Filipe, Luís (2018 web) *Machine-Assisted Proofs (ICM 2018 Panel)*, URL: <https://arxiv.org/pdf/1809.08062.pdf>
- Davies, Alex, Veličković, Petar, Buesing, Lars, Blackwell, Sam, Zheng, Daniel, Tomašev, Nenad, Tanburn, Richard, Battaglia, Peter, Blundell, Charles, Juhász, András, Lackenby, Marc, Williamson, Geordie, Hassabis, Demis, Kohli, Pushmeet (2021) “Advancing mathematics by guiding human intuition with AI”, *Nature*, Vol. 600, No. 7887, pp. 70–74.
- Denecker, Marc, Kakas, Antonis C. (2000) “Special issue: abductive logic programming”, *Journal of Logic Programming*, Vol. 44 (1–3), pp. 1–4.
- Dimitrov, Denis (2021 web) “*RuDALL-E: generating images from text descriptions, or the largest computational project in Russia*”, URL: <https://habr.com/ru/company/sberbank/blog/586926/>
- Dubrovsky, David I. (2010) *Deception. Philosophical and psychological analysis*, Canon +: Rehabilitation, Moscow (in Russian).
- Engineering for Research Symposium (2020 web) *Engineering for Research Symposium. Third Edition. Towards the Logic of Scientific Discovery. Will AI ever win a Nobel Prize?*, URL: <https://www.thoughtworks.com/engineering-research-symposium>
- Feickert, Matthew, Nachman, Benjamin P. (2021 web) “A Living Review of Machine Learning for Particle Physics”, URL: <https://arxiv.org/pdf/2102.02770.pdf>
- Goh, Gabriel, Cammarata, Nick, Voss, Chelsea, Carter, Shan, Petrov, Michael, Schubert, Ludwig, Radford, Alec, Olah, Chris (2021 web) “Multimodal neurons in artificial neural networks”, *Distill*, Vol. 6 (3), URL: <https://distill.pub/2021/multimodal-neurons/>
- Gonthier, Georges, Asperti, Andrea, Avigad, Jeremy, Bertot, Yves, Cohen, Cyril, Garillot et al (2013) “A Machine-Checked Proof of the Odd Order Theorem”, *Interactive Theorem Proving*, 4th Conference on Interactive Theorem Proving, Rennes, France, pp. 163–179.
- Howard, Don A., Giovanelli, Marco (2019 web) “Einstein’s philosophy of science”, *The Stanford Encyclopedia of Philosophy*, URL: <https://plato.stanford.edu/entries/einstein-philsience/>
- Kitano, Hiroaki (2016 web) “Nobel Turing Challenge: creating the engine for scientific discovery”, *Systems Biology and Applications*, Vol. 7, No. 1, URL: <https://www.nature.com/articles/s41540-021-00189-3#citeas>
- Larson, Erik J. (2021) *The Myth of Artificial Intelligence. Why Computers Can’t Think the Way We Do*, Harvard University Press, Cambridge.
- Leeming, Jack (2021) “How AI is helping the natural sciences”, *Nature*, Vol. 598, No. 7880, pp. 5–7.
- Ornes, Stephen (2021 web) “Mathematiker aus Silizium”, URL: <https://www.spektrum.de/news/koennen-computer-mathematische-beweise-fuehren/1815512> (in German).
- Ponce, Carlos R., Xiao, Will, Schade, Peter F., Hartmann, Till S., Kreiman, Gabriel, Livingstone, Margaret S. (2019) “Evolving Images for Visual Neurons Using a Deep Generative Network Reveals Coding Principles and Neuronal Preferences”, *Cell*, Vol. 177, No. 4, pp. 999–1009.
- Russell, Stuart, Norvig, Peter (2021) *Artificial Intelligence: A Modern Approach*, 4th US ed., Pearson, Hoboken.
- Savage, Neil (2019) “How AI and neuroscience drive each other forwards”, *Nature*, Vol. 571, No. 7766, pp. 15–17.
- Schaeffer, Jonathan (2007 web) *The Evolution of Knowledge and Search in Game-Playing Systems*, URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.116.9221&rep=rep1&type=pdf>
- Skrynnikov, Arseny (2021 web) “*Fantastically interesting properties*”: Russian scientist on new materials and technologies, URL: <https://ru.rt.com/hfg3>
- Smirnov, Stanislav K. (2021 web) *Conversation with a mathematician. Public interview with Stanislav Smirnov*, URL: <https://youtu.be/lvFu6IMXwj0>
- Sriraman, Bharath, ed. (2017) *Humanizing Mathematics and its Philosophy*, Springer International Publishing, Cham.
- Vardi, Moshe Y. (2021 web) “From Aristotle to the iPhone”, *World Logic Day Webinar*, URL: <https://www.youtube.com/watch?v=wOQuW6QFdos>

Сведения об авторах

АНОХИН Константин Владимирович – академик РАН, директор Института перспективных исследований мозга.

НОВОСЕЛОВ Константин Сергеевич – PhD в физике, профессор, лауреат Нобелевской премии по физике, Институт интеллектуальных функциональных материалов Национального университета Сингапура.

СМИРНОВ Станислав Константинович – PhD в математике, профессор, лауреат Филдсовской премии, профессор секции математики Женевского университета; научный руководитель Исследовательской лаборатории им. П.Л. Чебышева Санкт-Петербургского государственного университета; профессор Центра перспективных исследований Сколковского института науки и технологий.

ЕФИМОВ Альберт Рувимович – кандидат философских наук, директор Управления исследований и инноваций ПАО «Сбербанк», заведующий кафедрой инженерной кибернетики НИТУ «МИСиС».

МАТВЕЕВ Филипп Михайлович – студент философского факультета МГУ им. М.В. Ломоносова.

Author's Information

ANOKHIN Konstantin V. – Academician of the Russian Academy of Sciences, Institute for Advanced Brain Research.

NOVOSELOV Konstantin S. – PhD in physics, Professor, Nobel Prize winning physicist. Institute for Functional Intelligent Materials, National University of Singapore.

SMIRNOV Stanislav K. – PhD in Mathematics, Professor, Fields Prize winner, University of Geneva; academic supervisor, Chebyshev Laboratory, Saint Petersburg State University; full professor, Center for Advanced Studies, Skolkovo Institute of Science and Technology.

EFIMOV Albert R. – PhD in Philosophy, Director of the Department of Research and Innovation of PJSC Sberbank, Head of the Department of Engineering Cybernetics NUST "MISIS".

MATVEEV Philipp M. – student of the Faculty of Philosophy of Lomonosov Moscow State University.