

Нобелевские премии в 2024 получили группы ученых, которые разрабатывали ИИ-модели

Академик Олег Фиговский.

В 2024 году сразу две Нобелевские премии получили группы ученых, которые разрабатывали ИИ-модели - <https://n-n-n.ru/posts/ot-osnov-neirosetei-do-dizainerskix-belkov-ii-v-nobelevskix-premiiax-2024-goda>.

Эти награды очень разные — но обе признают фундаментальный вклад, который ИИ приносит в науку. При этом они очень четко обозначают место нейросетей в науке — это прежде всего ловкие и быстрые инструменты, который в руках талантливых ученых помогут построить лучший мир.

Нобелевская премия по физике 2024 года была присуждена Джону Хопфилду и Джеффри Хинтону с формулировкой «за основополагающие открытия и изобретения, которые позволяют осуществлять машинное обучение с использованием искусственных нейронных сетей».

Теоретические открытия этих двух ученых в той или иной степени предопределили все современные ИИ-решения, поэтому в заслуженности премии им не откажешь. Тем не менее, к физике как таковой они имеют достаточно опосредованное отношение.

Все началось в 1982 году, когда Джон Хопфилд, доктор физики Принстона с большим опытом биологических исследований, придумал первую модель рекуррентной нейронной сети. С ее помощью Хопфилд хотел с помощью физического инструментария изучить процессы человеческого мышления — и вряд ли мог представить, что его теоретическая модель ляжет в основу автоматических переводчиков или систем распознавания лиц. Сеть Хопфилда, как ее назвали в дальнейшем, описывала систему взаимодействующих между собой нейронов как физических объектов, между которыми возможны физические явления. Каждый нейрон в ней связан со всеми остальными и может как отправлять сигналы другим нейронам, так и получать их. Связи между нейронами характеризуются весами. Сеть хранит информацию в виде весов связей между нейронами. Эта информация может быть сгруппирована в несколько паттернов. Если в сеть поступает информация, напоминающая тот или иной паттерн, сеть может воспроизвести ее — даже если эта информация неполная или искаженная.

Воспроизведение шаблонов работает через функцию энергии. Для каждого из запомненных нейросетью шаблонов устанавливаются энергетические веса. Чем ближе сеть к воспроизведению того или иного шаблона, тем ниже ее общая энергия. Когда нейросеть получает входные данные, она вычисляет текущее состояние и соответствующую функцию энергии. Нейроны обновляют состояние, основываясь на весах связей и текущем состоянии, стремясь минимизировать энергию. Таким образом, нейросеть впервые получила ассоциативную

память, способную сохранять и воспроизводить шаблоны. А еще она получила способность учиться на примерах и создавать схожие с ними сущности по аналогии. В конечном счете это очень похоже на то, как учится человек. Этот принцип, который бесконечно развивали и усложняли многие ученые, лежит в основе современных рекуррентных нейросетей. Такую архитектуру, к примеру, используют нейросети для распознавания речи, изображений или видео и перевода, а также многие предсказательные и аналитические ИИ.

Исследования Джеффри Хинтона тоже оказались краеугольными для современных ИИ. Опираясь на работы своего коллеги и

соавтора Хопфилда, в середине 1980-х он разработал Больцмановскую машину (названа в честь австрийского физика Людвиг Больцмана) — модель нейронной сети, которая предвосхитила глубокое обучение. Машина Больцмана похожа на сеть Хинтона — в ней тоже работает энергетическая функция, сеть стремится к минимизации энергии. Однако она имеет стохастический, то есть вероятностный характер. Нейроны могут быть активными или неактивными, и их состояние определяется с помощью вероятностных распределений. Чем сильнее связаны два нейрона, тем выше шанс, что, если один из них активен, то второй тоже активируется.

Когда в машину Больцмана поступают данные, веса либо случайны, либо настроены человеком. Во время обучения она анализирует данные, изменяя свои веса так, чтобы более вероятные (повторяющиеся или закономерные) связи усиливались, а менее вероятные ослаблялись. Это позволяет нейросети находить схожести в данных без подсказок. К примеру, если обучить ее на множестве фотографий городских улиц, она заметит закономерности изображения автомобилей, зданий или людей, и запомнит их. После такого обучения, она сможет «узнавать» подобные объекты и на других фотографиях.

Открытия и принципы машины Больцмана, многократно углубленные и усложненные, используются в глубинных нейронных сетях, генеративных моделях, рекомендательных ИИ, обработке естественного языка — то есть в большинстве современных ИИ-решений.

Нобелевка Хинтона и Хопфилда выглядит прежде всего как признание мировым научным сообществом пользы мудрой и неординарной междисциплинарности. Попытавшись применить физический инструментарий к нейробиологии, они положили начало огромному прорыву, который повлиял на все науки. Едва ли сегодня найдется крупный научный прорыв, за которым не стоит смелого выхода за пределы традиционных границ научных дисциплин. А в науке эпохи развитого ИИ эта грань может стать совсем прозрачной.

Еще одна нобелевка этого года, связанная с искусственным интеллектом — по химии. Ее получили Дэвид Бейкер (50%) с мотивировкой «за предсказание структуры белков», и Демис Хассабис и Джон Джампер (50% вместе) за «за вычислительный дизайн белков». Эти немного размытые формулировки скрывают за

собой один из крупнейших прорывов в современной биологии — возможность создавать рукотворные белки.

Белки — основной строительный материал для живых организмов. Они состоят из длинных цепочек из 20 аминокислот. Микровзаимодействия между атомами в составе этих аминокислот определяют функции и поведение белка. В сумме они формируют его фолдинг, или сворачивание.

Свернутый белок выглядит как хаотично смятая веревка, но на самом деле это выверенная структура. Если белок свернут правильно, его внутренняя энергия минимизирована (прямо как у вспомнившей шаблон сети Хопфилда), и он выполняет свою функцию. Но если в свертывании была ошибка, то это нарушает работу клеток и приводит к их слипанию и гибели. Ошибки в свертывании могут происходить из-за мутаций генов, кодирующих белок, стрессовых факторов, воздействующих на клетку, а также естественного старения. Неправильное сворачивание белков — причина многих заболеваний, прежде всего дегенеративных: остеоартрита, болезни Альцгеймера или Паркинсона, диабета.

Ряд научных открытий второй половины XX века помог разобраться в том, как устроены белки и как они сворачиваются. В 1970-х стало понятно, что жизнь белков подчиняется очень сложной, многомерной и запутанной, но при этом строгой логике — и в теории можно распутать с помощью вычислений. От этого оставалась лишь пара шагов до возможности создавать новые белки, способные выполнять разные задачи. И эти шаги помогли сделать Бейкер, Хассабис и Джампер — а также ИИ.

Главная проблема белков в том, что количество возможных комбинаций аминокислот для белка длиной всего в 100 аминокислот превышает число атомов во вселенной. При этом самый длинный белок — титин, необходимый для сокращения скелетных мышц, состоит из 34 тысяч аминокислот. Но подавляющее большинство случайных последовательностей белков вообще не сворачиваются, то есть нежизнеспособны.

Дэвид Бейкер из Университета Вашингтона начал разрабатывать компьютерные программы для предсказания структур белка еще в 1990-х. Сперва они предназначались для того, чтобы предсказывать структуру и функции белков исходя из их аминокислотного строения. Но оказалось, что этот метод можно запустить наоборот — сначала определить функцию белка, потом — его структуру, и затем подобрать его аминокислотное строение. Это в теории позволяло конструировать белки *ab initio* («с нуля»). Этот подход очень сильно отличался от общепринятого метода, в котором ученые оттачивались от сходств в строении белков. Но, как говорил Бейкер, «если вы хотите построить самолет, вы не начинаете с модификации птицы — вместо этого вы понимаете основные принципы аэродинамики и строите летательные аппараты на основе этих принципов». На рубеже 1990-х и 2000-х команда Бейкера разработала вычислительную методику на FORTRAN, которая получила название Rosetta (названа в честь камня, который помог расшифровать древнеегипетское письмо). Она опиралась на массив экспериментальных данных, собранных

лабораторией Бейкера.

Алгоритм совершил прорыв в предсказании структуры белков. В 2001 году Бейкер и его Rosetta с разгромным счетом победили в конкурсе «Критическая оценка предсказания структуры белка» (CASP). Этот конкурс — что-то вроде главного чемпионата для биологов, которые изучают структуры белков. Раз в два года группы исследователей со всего мира получали последовательности аминокислот, которые не исследовались ранее, и должны были предсказать структуру белка на их основе. Команда Бейкера набрала 31 очко, а группа, ставшая серебряным призером — всего 8 очков.

А в 2003 году Бейкеру удалось с помощью Rosetta создать стабильный искусственный белок — Top7, длиной в 93 аминокислоты. Он не был первым рукотворным белком — еще в 1997 году в Колтехе создали белок FSD-1 из 28 аминокислот. Но если FSD-1 был создан по шаблону природного белка, то Top7 — разработан с нуля, и намеренно сделан так, чтобы быть непохожим ни на один, обнаруженный в природе.

Rosetta начала стремительно развиваться. Главной проблемой стало отсутствие вычислительных мощностей — в 2005 году ученые запустили проект Rosetta@home, который позволяет любому человеку предоставить ученым вычислительные мощности своего компьютера. В 2008 году они создали белки-ферменты, которые в отличие от Top7, который не выполнял никакой функции, проявляли функциональную активность. Сегодня новые белки для фармацевтики, химической промышленности и исследовательских целей создаются повсюду. Лаборатория Бейкера и его проекты создали несколько тысяч новых белков, которые используются в противораковых и противовирусных препаратах, вакцинах — в том числе и в вакцине от коронавируса от Университета Вашингтона. Кстати, Бейкер — это не только талантливый ученый, но и серийный предприниматель. Ему принадлежит более 100 патентов. Кроме того, он участвовал в создании 21 биотехнологической компании. Стартапы, к которым приложил руку Бейкер, покупают фармгиганты.

Примеры патентов Дэвида Бейкера в последнее время приводятся ниже:

US 20240368233 - MODULAR RECONFIGURABLE ASYMMETRIC PROTEIN ASSEMBLIES - Polypeptides and fusion proteins capable of heterodimer formation, methods for their use, and methods for their design are provided.

US 20240337655 - SILK FIBROIN TO STABILIZE "DE NOVO" SENSING PROTEINS - A solid form sensing article can have one or more silk fibroin matrix sensing units, each having a solid form sensing property. Each of the one or more silk fibroin matrix sensing units can include a silk fibroin matrix. A plurality of Cage proteins and a plurality of Key proteins can be entrained within the silk fibroin matrix. The Cage proteins can bind to an analyte of interest.

US 20240330569 - SYSTEM AND METHOD FOR XBRL REVIEW -

Provided is a method for review of an XBRL document. An XBRL document is received followed by a request to create an XBRL review for the XBRL document. A selection of an XBRL fact included in the XBRL document is received and a first review status, of a plurality of review statuses, is applied to the selected XBRL fact. The selected XBRL fact is monitored to detect a change to the XBRL fact, and in response to detecting a change to the XBRL fact, the first review status of the XBRL fact is updated to a second review status of the plurality of review statuses.

US 20240263167 - HIGH THROUGHPUT PROTEIN-PROTEIN INTERACTION SCREENING IN YEAST LIQUID CULTURE - The present invention relates to methods and compositions for the high throughput screening of protein-protein interactions in yeast liquid culture. Protein fusions non-native to yeast may be expressed to replace endogenous sexual agglutination proteins and mediate library-by-library interrogation of protein interactions. The methods and compositions of the invention can be utilized for the characterization of protein interaction networks in high throughput for both binding affinity and specificity, which is crucial for understanding cellular functions, screening therapeutic candidates, and evaluating engineered protein networks.

US 20240207376 - Compositions and Methods for Treating Celiac Sprue Disease - The invention provides compositions and methods for treating celiac sprue.

US 20240165218 - SELF-ASSEMBLING NANOSTRUCTURE VACCINES - The present disclosure provides nanostructures and nanostructure-based vaccines. Some nanostructures of the present disclosure display antigens capable of eliciting immune responses to infectious agents such as bacteria, viruses, and pathogens. Some vaccines of the present disclosure are useful for preventing or decreasing the severity of infection with an infectious agent, including, for example and without limitation, lyme disease, pertussis, herpes virus, orthomyxovirus, paramyxovirus, pneumovirus, filovirus, flavivirus, reovirus, retrovirus, meningococcus, or malaria. The antigens may be attached to the core of the nanostructure either non-covalently or covalently, including as a fusion protein or by other means disclosed herein. Multimeric antigens may optionally be displayed along a symmetry axis of the nanostructure. Also provided are proteins and nucleic acid molecules encoding such proteins, vaccine compositions, and methods of administration.

US 20240157596 - LOG SPLITTER ACTUATION ASSEMBLY - A log splitter is provided. The log splitter includes a housing, a power source, a motor within the housing and electrically coupled to the power source, a support arm coupled to and at least partially within the housing, a rack coupled to at least a portion of the support arm, and an actuation assembly at least partially within the housing and coupled to the motor and the rack. The power source may include at least one battery. The support arm extends between a first end and a second end. The rack is configured to move along the support arm in a forward direction to an extended position adjacent the first end of the support arm and a reverse direction toward to a starting position adjacent the second end of the support arm. The actuation assembly is configured to move the rack at least in the forward direction.

Так, Takeda Pharmaceuticals купила PVP Biologics за \$330 млн — стартап разработал белок-суперглютеназу, которым можно лечить непереносимость глютена. А в 2024 году AstraZeneca за \$1,1 млрд купила стартап Icosavax, который разрабатывает синтетические вакцины, нацеленные на вирусы естественного происхождения.

Хассабис и Джампер продолжили работу Бейкера, ускорив ее с помощью ИИ. Эти двое — довольно редкий случай, когда Нобелевку получают не академические, а корпоративные исследователи. Модель AlphaFold была создана в стенах DeepMind — ИИ-подразделения Google. Помните конкурс CASP, который в 2003 году с разгромным счетом выиграла Rosetta? Спустя 15 лет первое место в нем занял другой инструмент — нейросеть AlphaFold от Google DeepMind. Напомним, что Хассабис стал одним из сооснователей DeepMind в 2010 году. А четыре года спустя ее купил Google, оставив Хассабиса ее гендиректором. Первая версия AlphaFold была эффективна, но прорывом не стала. Точность ее прогнозов по метрикам CASP была выше, чем у конкурентов, но все еще была ниже золотого стандарта структурной биологии. К тому же другие участники конкурса тоже применяли ИИ и лишь немного отставали от DeepMind.

«Если вы посмотрите, у кого была лучшая модель для каждой из целей конкурса в 2018 году, то чаще всего это была DeepMind, чем кто-то еще, — рассказывал основатель конкурса CASP Джон Моулт. — Но и другие участники тоже оказывались лучшими. Так что это был сюрприз, но не такой уж и потрясающий». Все изменилось после того, как в 2018 году в компанию пришел Джон Джампер, вскоре возглавивший AlphaFold. Как и Хопфилд за три с половиной десятилетия до него, Джампер — физик-теоретик, который заинтересовался возможностями применения физических подходов в биологии.

Под руководством Джампера команда AlphaFold фундаментально переработала этот инструмент. AlphaFold первой версии использовала сверхточные нейронные сети (CNN), которые действовали примерно так же, как нейросети для распознавания изображения. Из-за этого они были ограничены в улавливании отдаленных взаимодействий между аминокислотами, особенно в больших белках. На первом этапе ИИ на основе генетических данных прогнозировал расстояние между парами аминокислот в белке. На втором этапе — уже без ИИ, с помощью этой информации AlphaFold разрабатывает консенсусную модель белка. Но AlphaFold2 работает на основе архитектуры с использованием трансформеров. Нейросеть анализирует эволюционно близкие белковые последовательности и выявляет закономерности и связи между аминокислотами, которые важны для формирования структуры их свертки. А трансформеры позволяют находить не только локальные, но и отдаленные связи между аминокислотами, что существенно повышает точность моделирования сложных белков.

На основании данных и предсказанных ограничений, AlphaFold строит вероятные промежуточные модели белка, которые затем оптимизируются до получения наиболее вероятной структуры с

минимальной энергией. После этого нейросеть оценивает достоверность моделей белков с использованием встроенной метрики, которая позволяет ей отбрасывать менее точные варианты и сосредотачиваться на структурах, соответствующих реальности.

Позднее выяснилось, что AlphaFold2 требуется лишь 1% данных для обучения от того объема, что нужен первой версии, чтобы достичь лучших результатов. На CASP 2020 года AlphaFold2 вновь победила — но на этот раз с гораздо большим отрывом от остальных участников, включая команды крупнейших компаний и исследовательских центров, которые тоже вовсю использовали ИИ. Точность цифрового метода предсказания впервые достигла результативности, сравнимой с экспериментальными методами. Уже в 2022 году нейросеть смогла предсказать трехмерные структуры 200 миллионов белков — почти всех, известных науке, и опубликовала данные о них в открытом доступе. То, что раньше требовало бы нескольких месяцев работы, AlphaFold выполняла за несколько секунд.

Возможности AlphaFold используются в сотнях и тысячах исследовательских и фармацевтических проектов — от вакцины от малярии с белком Pfs48/45, который может сделать ее гораздо эффективнее, работать на всех стадиях болезни и блокировать ее передачу, до потенциальных методов защиты пчел, лягушек и крокодилов от негативного влияния пестицидов. В мае этого года Google DeepMind представил AlphaFold3 — но ее функционал доступен для исследователей бесплатно в меньшей степени, чем для предыдущей модели, а эксперты говорят, что она не произведет такую революцию, как ее предшественница.

Две эти Нобелевские премии красноречиво показывают, как стремительно ИИ меняет мир. Когда речь заходит о подрывном потенциале нейросетей, большинство вспоминает о Midjourney, который оставит без работы дизайнеров, или о страшных рассказах «ИИ-думеров» о том, как общий ИИ поработает или уничтожает человечество. Но по крайней мере пока подрывные (в экономическом смысле) изменения совсем иные — научные открытия, которые в течение одного-двух десятилетий могут радикально улучшить жизнь миллионов людей.

В заключение приведем некоторые патенты, базирующиеся на работах лауреатов Нобелевских премий.

UA77519 (U) - METHOD FOR PROGRAMMING ARTIFICIAL INTELLECT BY MEANS OF TERMINALS OF PARTICIPANTS OF GEOSOCIAL GAMES, INTERNET AND MOBILE CONNECTION WITH ACCOUNT OF CLAIMED RESULTING COLLECTIVE RESPONSE - A method for programming artificial intellect by means of terminals of participants of geosocial games, Internet and mobile connection with account of claimed resulting collective response is performed by transmitting text, graphic, audio, video responses or response as program code by sending SMS, MMS, vocal, tone and USSD communications, or Internet-communication by means of subscriber terminals, with further allocation of information on Internet-server so as

to be represented at Internet-site. Answers to the questions received from participants by means of two-way communications along with data on geographical location of gamers are collected, stored, analyzed and processed on special server by means of corresponding programs, simultaneously if necessary by means of corresponding interfaces to program complex, they are independently accounted, controlled, analyzed with the help of operators, and thereupon are sent to server of data processing as one or several resulting responses for further programming of artificial intellect.

EP1691665 (A2) - ARTIFICIAL INTELLECT AND DEVICE FOR DIAGNOSTICS, SCREENING, PREVENTION AND TREATMENT OF STATES OF MOTHER-FETUS SYSTEM - FIELD: medicine. ^

SUBSTANCE: system for health optimisation during pregnancy contains database with data about connected with pregnancy states of health, input device for input of diagnostic and/or screening patient's data, including oriented in time information about said diagnostic and screening data, indicator device for displaying report about decision depending on fed diagnostic and screening data and said data about connected with pregnancy states of health, and intellectual agent. Database additionally contains oriented in time information about complications of health state during pregnancy. Intellectual agent contains algorithmic rule developed for appealing to data introduced into intellectual agent. Intellectual agent applies algorithmic rule for producing decision about pregnancy and uses said device to form report about said decision. ^ EFFECT: application of invention will allow to increase probability of newborn appearing in the best state for entering life.

KR20010047140 (A) - ARTIFICIAL INTELLECT TYPE EYE-PROTECTIVE WINDOW OF HELMET FOR RIDING - PURPOSE: An artificial intellect type eye-protective window of helmet for riding is provided to prevent window being clouded up with steam even at a low temperature and by coating surface lens with transparent thermal conductive materials and supplying powder. CONSTITUTION: The helmet protective window comprises exposure lens(3) exposed to the outside, surface lens(4) in the inside of face which is coated with transparent thermal conductive metal film, an air gap(5) between exposure lens and surface lens, a thermal line(2) silk-printed on the surface lens and a controller supplying power based on the temperature checked by thermal sensor, namely, comparing to the standard value. The helmet protective window is used at a range of -25 to 40deg.C and prevents from being overheated by thermal sensor.

Будем ждать лауреатов Нобелевских премий 2025 года и новых приоритетных направлений в науке.