

Новиков Н.Б.

Институт психологии РАН

Россия, г. Москва

Novikov N.B.

Institute of Psychology RAS

Russia, Moscow

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ДОЛЖЕН НАУЧИТЬСЯ ДЕЛАТЬ СЛУЧАЙНЫЕ ОТКРЫТИЯ

Аннотация: В 1989 г. британский физик и математик Роджер Пенроуз опубликовал книгу «Новый ум короля», в которой продемонстрировал, что теорема Геделя о неполноте, а также теорема Тьюринга о неразрешимости проблемы останковки являются препятствиями для полной формализации (алгоритмизации) деятельности интеллекта: как человеческого, так и искусственного. В 1997 г. американский математик Стивен Смейл, составляя свой список нерешенных математических проблем, интерпретировал упомянутые теоремы Геделя и Тьюринга как пределы интеллекта. 18-я проблема С.Смейла звучит так: каковы пределы интеллекта: как искусственного, так и человека? В 2013 г. мы показали, что пределами формализации интеллекта, помимо теорем Геделя и Тьюринга, являются три важных фактора: вероятностная природа индуктивного вывода, метод проб и ошибок (метод последовательного перебора) и фактор случая в научном открытии. Фактор случая заключается в том, что имеется широкий класс научных открытий, которые являются непреднамеренными (незапланированными, случайными). Мы не можем предсказывать такие находки, но они вносят существенный вклад в прогресс науки. Учитывая это

обстоятельство, а также тот факт, что специалисты в области компьютерных наук пытаются создать искусственный интеллект, способный делать научные открытия, мы формулируем следующий вывод: нам следует изучить механизм случайных открытий, чтобы наделить системы искусственного интеллекта (которые появятся в будущем) способностью делать подобные открытия.

Ключевые слова: теорема Геделя о неполноте, теорема Тьюринга о неразрешимости проблемы остановки, 18-я проблема С.Смейла, случайные открытия, искусственный интеллект.

Abstract: In 1989, the British physicist and mathematician Roger Penrose published the book “The Emperor’s New Mind”, in which he demonstrated that Gödel’s incompleteness theorem, as well as Turing’s theorem on the unsolvability of the halting problem, are obstacles to the complete formalization (algorithmization) of the activity of the intellect: both human, as well as artificial. In 1997, the American mathematician Stephen Smale, compiling his list of unsolved mathematical problems, interpreted the mentioned theorems of Gödel and Turing as the limits of intelligence. The 18th problem of S. Smale sounds like this: what are the limits of intelligence: both artificial and human? In 2013, we showed that the limits of the formalization of intelligence, in addition to the Gödel and Turing theorems, are three important factors: the probabilistic nature of inductive inference, the trial and error method (successive elimination method), and the chance factor in scientific discovery. The chance factor is that there is a wide class of scientific discoveries that are unintentional (unplanned, accidental). We cannot predict such findings, but they make a significant contribution to the progress of science. Given this circumstance, and the fact that researchers are trying to create artificial intelligence capable of making scientific discoveries, we formulate the following conclusion: we should study the mechanism of random discoveries in order to endow artificial intelligence systems (which will appear in the future) with the ability to make such discoveries.

Keywords: Godel's incompleteness theorem, Turing's theorem on the undecidability of the halting problem, 18th problem of S. Smale, random discoveries, artificial intelligence.

1. О теоремах Геделя и Тьюринга

В 1931 г. австрийский математик Курт Гедель сообщил о том, что ему удалось доказать теорему, согласно которой в любой формальной системе типа арифметики можно сформулировать утверждения, истинность которых нельзя установить средствами самой арифметики. Его доказательство было проверено множеством специалистов, которые убедились в справедливости аргументов (математических выкладок), опубликованных молодым австрийским ученым. Основной результат работы К.Геделя получил название «теоремы Геделя о неполноте». Стало ясно, что эта теорема ставит под сомнение осуществимость программы формализации математики, предложенной Давидом Гильбертом (1862-1943). Один из ассистентов Д.Гильберта, венгерский математик Джон фон Нейман, ознакомившись с результатом К.Геделя, сделал вывод, что настало время менять область математических исследований. Фон Нейман прекратил работать в рамках программы формализации Д.Гильберта, переключившись на решение других научных проблем.

В 1936 г. британский математик Алан Тьюринг доказал еще одну теорему, продемонстрировавшую неосуществимость программы Д.Гильберта, - теорему о неразрешимости проблемы остановки. Вскоре среди специалистов стала распространяться точка зрения о том, что, если мы не можем достичь полной формализации математики, значит, нам не удастся достичь полной формализации человеческого интеллекта. Эта мысль основывалась на следующем простом рассуждении: теоремы Геделя и Тьюринга налагают ограничения на возможность формализации математики, в том числе

математического творчества. Поскольку математическое творчество – одно из проявлений (одна из сторон) человеческого творчества, то следует признать невозможность полного формального описания творческой деятельности человеческого интеллекта.

В 1946 г. Герман Вейль (который, как и фон Нейман, был когда-то ассистентом Д.Гильберта) опубликовал в журнале «Американский математический ежемесячник» статью под названием «Математика и логика». В ней он рассмотрел ключевые моменты программы формализации, которой до 1931 г. занимались многие ученые, воодушевленные идеями Д.Гильберта, и отметил, что теорема Геделя о неполноте стала причиной переоценки прежних ценностей. В частности, Г.Вейль в указанной статье [1] пишет: «Видимо, все математики вынуждены были бы, в конце концов, признать, что подход Гильберта может успешно привести к намеченной цели. Первые шаги были вдохновляющими и многообещающими. Но тут Гедель нанес этой точке зрения сокрушительный удар, от которого она не оправилась до сих пор. Гедель определенным образом перенумеровал символы, формулы и последовательности формул формализма Гильберта, а затем преобразовал утверждение о непротиворечивости в арифметическое высказывание. Он показал, что это высказывание не может быть ни доказано, ни опровергнуто внутри формализма» [1, с.339].

В 1951 г. Г.Вейль опубликовал в том же «Американском математическом ежемесячнике» еще одну статью, в которой обсуждалась теорема Геделя. Эта статья была издана в СССР в виде брошюры под названием «Полвека математики» [2]. В данной работе, посвященной успехам математики за первые пять десятилетий XX века, Г.Вейль говорит о Д.Гильберте: «Его попытки доказать непротиворечивость обнаружили поразительно сложную логическую структуру математики. Первые шаги были действительно обнадеживающими. Но затем открытие Геделя бросило глубокую тень на гильбертово предприятие. Непротиворечивость сама может быть выражена

формулой. Гедель показал следующее: если игра, в которую играют математики, действительно непротиворечива, то формула непротиворечивости в этой игре вообще не может быть доказана. Но если дело обстоит таким образом, то как же мы можем вообще надеяться доказать эту формулу?» [2, с.44].

После доказательства теоремы о неразрешимости проблемы остановки А.Тьюринг занялся разработкой электронных вычислительных машин. В период времени с 1939 по 1945 гг. он участвовал в проекте по созданию машин, способных расшифровывать секретные коды. Совместно с коллегами ему удалось построить реальную электромеханическую машину, включавшую два фотоэлектрических устройства считывания информации с перфоленты со скоростью 2000 символов в секунду, арифметическое устройство на реле и печатающий блок. Эта машина работала по принципу перебора различных комбинаций из символов секретного кода, который подлежал дешифровке. В 1943 г. А.Тьюринг и его сотрудники разработали уже электронную (а не электромеханическую) вычислительную машину, названную «Колосс», которая содержала 2000 электронных ламп. В это время А.Тьюринг совершил поездку в США, где встречался с Джоном фон Нейманом (одним из ассистентов Д.Гильберта) и делился с ним идеями о возможностях вычислительных машин. Общение с А.Тьюрингом, а также работа над созданием первой атомной бомбы в Лос-Аламосе (США) убедили фон Неймана в высокой значимости компьютеров. Вскоре он разработал проект универсальной вычислительной машины, способной быстро извлекать из памяти (хранимой программы) необходимую информацию и обрабатывать ее. Это стало началом проектирования самых разных вычислительных машин, ввиду чего фон Неймана (наряду с Тьюрингом) называют отцом этих устройств [3].

Появление компьютеров, выполнявших множество интеллектуальных функций (ранее доступных лишь человеку), позволяло расширить сферу

действия теоремы Геделя о неполноте. Ученые знали, что вычислительные машины, решающие те или иные арифметические задачи, действуют на основе системы предписаний, аналогичных тем, что изложены Д.Гильбертом и его учениками в попытках формализовать арифметику (и всю математику). Отсюда возникал вывод, что ограничения, связанные с теоремой Геделя, должны распространяться также на компьютеры, чьи программы основаны на использовании формальных (замкнутых) алгоритмов.

Впервые этот вывод сформулировали американские математики Эрнест Нагель и Джеймс Ньюмен в книге «Доказательство Геделя» (1958). Их книга была издана в СССР в 1970 г. под названием «Теорема Геделя». В ней авторы аргументируют [4]: «Заключения, к которым пришел Гедель, порождают, естественно, и вопрос, можно ли построить вычислительную машину, сравнимую по своим «творческим» математическим возможностям с человеческим мозгом. Современные вычислительные машины обладают некоторым точно фиксированным запасом команд, которые умеют выполнять их элементы и блоки; команды соответствуют фиксированным правилам вывода некоторой формализованной аксиоматической процедуры. Таким образом, машина решает задачу, шаг за шагом выполняя одну из «встроенных» в нее заранее команд. Однако, как видно из геделевской теоремы о неполноте, уже в элементарной арифметике натуральных чисел возникает бесчисленное множество проблем, выходящих за пределы возможностей любой конкретной аксиоматической системы, а значит, и недоступных для таких машин, сколь бы остроумными и сложными ни были их конструкции и с какой бы громадной скоростью ни проделывали они свои операции. Для каждой конкретной задачи в принципе можно построить машину, которой эта задача была бы под силу, но нельзя создать машину, пригодную для решения любой задачи. Правда, и возможности человеческого мозга могут оказаться ограниченными, так что и человек тогда сможет решить не любую задачу» [4, с.58-59].

В 1975 г. этот вывод повторил отечественный математик Юрий Иванович Манин (1937-2023). В статье «Теорема Геделя» [5] он отметил, что результат Геделя налагает ограничения на дедуктивные способы рассуждений, поскольку именно эти рассуждения лежат в основе формальных аксиоматико-дедуктивных систем. Ю.И.Манин пишет: «...Мы знаем, что метод дедуктивных выводов недостаточно мощен. Его не хватает даже на то, чтобы вывести из конечного числа принципов все истинные утверждения о целых числах, формулируемые на языке школьной алгебры: таков смысл теоремы Геделя. Это осознание глубокой ограниченности дедуктивных рассуждений и вообще «механических» методов поиска истины стало особенно актуально в эпоху экспансии ЭВМ. Изложенная точка зрения на теорему Геделя дает основания считать ее существенным вкладом естественных наук в фонд гуманитарных» [5, с.80].

Таким образом, до публикации книги Р.Пенроуза «Новый ум короля» в науке уже существовало понимание ограничений, вытекающих из теоремы Геделя (а также результата Тьюринга). В частности, высказывалась точка зрения о том, что теорема Геделя о неполноте запрещает продуктивную творческую деятельность, ограниченную применением дедуктивных рассуждений (разновидности детерминированных алгоритмов). Причем не важно, кто является субъектом этой деятельности – человек или искусственный интеллект. Однако выводы, сделанные Э.Нагелем и Дж.Ньюменом в 1958 г., когда появлялись первые вычислительные машины, могли показаться не вполне справедливыми в 1989 г., когда компьютеру оставалось девять лет до победы над чемпионом мира по шахматам Г.Каспаровым.

Заслуга Р.Пенроуза (1989) заключается в том, что он подтвердил предположение своих предшественников о том, что любая машина, основанная на дедуктивной логике (любая дедуктивная ЭВМ), неизбежно столкнется с запретами, вытекающими из результатов Геделя и Тьюринга.

Кроме того, используя идеи теории вычислимости - относительно новой в математике теории, - Р.Пенроуз показал, что внутри формальной системы (внутри любой дедуктивной программы «думающей» машины) понятие истины невычислимо. Другими словами, поскольку творчество не является процессом, который можно описать рекурсивными функциями (эти функции обеспечивают вычислимость процесса), то оно, творчество, связанное с открытием новых математических и вообще научных истин, не относится к числу вычисляемых явлений. Отсюда следует заключение о невычислимости процесса научного открытия, о невозможности полностью формализовать (алгоритмизировать) этот процесс. В книге «Новый ум короля» [6] британский математик пишет: «...В нашей способности познавать - а, следовательно, и в нашей сознательной деятельности в целом – есть нечто, выходящее за пределы чисто алгоритмических действий...» [6, с.14]. «...Из доказательства Геделя следует с очевидностью, что понятие математической истины не может быть заключено ни в одну из формальных систем. Математическая истина выходит за рамки любого формализма» [6, с.101].

Внимательно проанализировав эти аргументы Р.Пенроуза (который, кстати, в 2020 г. удостоен Нобелевской премии по физике), американский математик Стивен Смейл интерпретировал теоремы Геделя и Тьюринга как пределы интеллекта: человеческого и искусственного. Правильнее было бы сказать: пределы формализации (алгоритмизации) интеллекта. В 1997 г., составляя список нерешенных математических проблем, С.Смейл сформулировал 18-ю проблему следующим образом: каковы пределы интеллекта – как искусственного, так и человека? Иначе говоря, американский математик догадывался, что, помимо теоремы Геделя о неполноте и теоремы Тьюринга о неразрешимости проблемы остановки, должны существовать другие факторы, препятствующие формализации интеллекта.

Нам посчастливилось открыть эти факторы (пределы формализации творческой деятельности). В 2013 г. мы [7], [8] показали, что существует, по

меньшей мере, еще три аспекта научного творчества, которые не позволяют нам полностью алгоритмизировать процесс поиска истины (процесс познания природы). Первый из них – вероятностная природа индуктивного вывода, второй – метод проб и ошибок (который также содержит элемент стохастичности, неопределенности) и, наконец, третий аспект – фактор случая в научном открытии. Этот фактор означает, что существует широкий класс научных открытий, которые являются непреднамеренными (незапланированными, случайными). Случайные открытия, как отмечено выше, не поддаются точному предсказанию, но вносят значительный вклад в прогресс науки.

Нужно отметить, что фактор случая в научном открытии – более фундаментальное ограничение (препятствие) для формализации интеллекта, т.е. научного творчества, чем теоремы Геделя и Тьюринга. Дело в том, что запреты, связанные с этими теоремами, снимаются, если мы наделяем вычислительные машины способностью заимствовать информацию, необходимую для тех или иных целей, из внешнего мира. Машины, способные черпать сведения из наблюдения и эксперимента (как это делает человек), приобретают статус открытых систем. В этом случае к ним неприменимы ограничения, вытекающие из теорем Геделя и Тьюринга, которые относятся, как мы знаем, к замкнутым формальным системам, не использующим результатов эксперимента и наблюдения для их последующего обобщения. Что касается фактора случая в научном поиске, то его нельзя устранить в ситуации, когда вычислительные машины приобретают статус открытых систем. Иначе говоря, статус открытых систем не упраздняет неопределенность (стохастичность) указанного фактора случая, не устраняет условий, в которых совершаются случайные научные открытия. В этом, пожалуй, состоит одно из обстоятельств, заставляющих внимательно исследовать механизм случайных открытий.

2. Роль случайных открытий в прогрессе науки

Многие математики, теоретики искусственного интеллекта, а также инженеры, создающие и совершенствующие технологии машинного обучения, не интересуются случайными открытиями и не изучают обстоятельства, которые определяют (обуславливают) эти открытия. Собственно говоря, ничего удивительного в этом нет: упомянутые математики и инженеры заняты множеством других (специфичных для их деятельности) проблем, и им некогда анализировать феномен непреднамеренных находок. Этот феномен должны изучать науковеды, методологи науки и, прежде всего, историки научного знания, цель которых – правдиво и детально описать генезис каждого крупного научного результата. И можно сказать, что они замечательным образом справились с этой задачей! Этому в значительной степени способствовали и сами первооткрыватели, которые сообщали в своих публикациях (в том числе воспоминаниях) о том, как им удавалось обнаруживать нечто новое. Если ограничиться рассмотрением ученых, чьи научные достижения удостоены Нобелевской премии, то и в этом случае мы получим достаточно большой список открытий, сделанных по принципу «серендипити» (serendipity): искал одно, нашел другое.

Теперь мы знаем, что фактор случая сыграл свою роль в следующих открытиях, определивших современную научную картину мира:

- открытие рентгеновских лучей (Вильгельм Рентген, Нобелевская премия по физике за 1901 г.);
- открытие радиоактивности (Анри Беккерель, Нобелевская премия по физике за 1903 г.);
- открытие сверхпроводимости (Хейке Камерлинг-Оннес, Нобелевская премия по физике за 1913 г.);
- открытие элементарной частицы, названной «позитроном» (Карл Андерсон, Нобелевская премия по физике за 1936 г.);

- открытие космических лучей (Виктор Гесс, Нобелевская премия по физике за 1936 г.);
- открытие дифракции электрона (Клинтон Дэвиссон, Нобелевская премия по физике за 1937 г.);
- открытие эффекта замедления нейтронов (Энрико Ферми, Нобелевская премия по физике за 1938 г.);
- открытие эффекта Черенкова-Вавилова (Павел Алексеевич Черенков, Нобелевская премия по физике за 1958 г.);
- открытие мюонного катализа (Луис Альварес, Нобелевская премия по физике за 1968 г.);
- открытие нейтронных звезд (Энтони Хьюиш, Нобелевская премия по физике за 1974 г.);
- открытие микроволнового космического реликтового излучения (Арно Пензиас, Роберт Вильсон, Нобелевская премия по физике за 1978 г.);
- открытие нарушения комбинированной симметрии в физике частиц (Джеймс Кронин, Вал Фитч, Нобелевская премия по физике за 1980 г.);
- открытие двойного пульсара PSR 1513-10, позволившего косвенно подтвердить существование гравитационных волн (Рассел Халс, Джозеф Тейлор, Нобелевская премия по физике за 1993 г.);
- открытие способа культивирования микробов на твердых поверхностях (Роберт Кох, Нобелевская премия по физиологии и медицине за 1905 г.);
- открытие явления анафилаксии (Шарль Рише, Нобелевская премия по физиологии и медицине за 1913 г.);
- открытие витаминов (Христиан Эйкман, Нобелевская премия по физиологии и медицине за 1929 г.);
- открытие пенициллина (Александр Флеминг, Нобелевская премия по физиологии и медицине за 1945 г.);
- открытие стероидного гормона кортизона (Филип Хенч, Нобелевская премия по физиологии и медицине за 1950 г.);

- открытие свойства полиурацила инициировать синтез белка полифенилаланина, что позволило расшифровать генетический код (Маршалл Ниренберг, Нобелевская премия по физиологии и медицине за 1968 г.);
- открытие простагландинов (Ульф фон Эйлер, Нобелевская премия по физиологии и медицине за 1970 г.);
- открытие лизосом – клеточных органелл (Кристиан де Дюв, Нобелевская премия по физиологии и медицине за 1974 г.);
- открытие австралийского антигена и вируса гепатита В (Барух Бламберг, Нобелевская премия по физиологии и медицине за 1976 г.);
- открытие фактора роста нервов (ФРН) в змеином яде (Стенли Коэн, Нобелевская премия по физиологии и медицине за 1986 г.);
- открытие биологической функции оксида азота (Роберт Ферчготт, Нобелевская премия по физиологии и медицине за 1998 г.);
- открытие РНК-интерференции (Эндрю Файер, Крэйг Мелло, Нобелевская премия по физиологии и медицине за 2006 г.);
- открытие инертных газов гелия и криптона (Вильям Рамзай, Нобелевская премия по химии за 1904 г.);
- открытие бесклеточного брожения, т.е. внеклеточной ферментации (Эдуард Бухнер, Нобелевская премия по химии за 1907 г.);
- открытие атомного ядра (Эрнест Резерфорд, Нобелевская премия по химии за 1908 г.);
- открытие катализаторов синтеза аммиака из азота (Карл Бош, Нобелевская премия по химии за 1931 г.);
- открытие электронной эмиссии на поверхности вольфрамовой нити (Ирвинг Ленгмюр, Нобелевская премия по химии за 1932 г.);
- открытие явления деления атома урана (Отто Хан, Нобелевская премия по химии за 1944 г.);
- открытие трансуранового элемента № 99, названного «эйнштейний» (Гленн Сиборг, Нобелевская премия по химии за 1951 г.);

- открытие фактов, позволивших построить теорию разветвленных цепных химических реакций (Николай Николаевич Семенов, Нобелевская премия по химии за 1956 г.);
- открытие катализаторов синтеза полиэтилена (Карл Циглер, Нобелевская премия по химии за 1963 г.);
- открытие реакции Виттига – реакции получения олефинов из карбонильных соединений (Георг Виттиг, Нобелевская премия по химии за 1979 г.);
- открытие краун-эфиров (Чарльз Педерсен, Нобелевская премия по химии за 1987 г.);
- открытие каталитических свойств молекулы РНК (Томас Чех, Нобелевская премия по химии за 1989 г.);
- открытие первого фуллерена – молекулы углерода, состоящей из 60 атомов (Роберт Керл, Харольд Крото, Ричард Смолли, Нобелевская премия по химии за 1996 г.);
- открытие электропроводящих полимеров (Хидеки Ширакава, Алан Хигер, Алан Мак-Диармид, Нобелевская премия по химии за 2000 г.);
- открытие молекулы – части древней иммунной системы микробов CRISPR/Cas9, позволившей разработать метод генетического редактирования (Эммануэль Шарпантье, Нобелевская премия по химии за 2020 г.);
- открытие азид-алкинового циклоприсоединения, катализируемого медью, что легло в основу «клик-химии» (Мортен Мельдаль, Нобелевская премия по химии за 2022 г.).

Подробное описание этих непреднамеренных научных находок можно найти в работе [9]. В данную работу, к сожалению, не вошли открытия Э.Шарпантье и М.Мельдаля, так как информация об истории их научных достижений стала известна лишь после выхода книги [9].

Среди исследователей, которые откровенно сообщали о непреднамеренности своих открытий, мы находим Вильгельма Рентгена,

обнаружившего рентгеновские лучи, Клинтона Дэвиссона, открывшего дифракцию электронов, Энрико Ферми, описавшего эффект замедления нейтронов, Энтони Хьюиша, зафиксировавшего первые сигналы нейтронных звезд (радиопульсаров). Здесь же следует упомянуть Джеймса Кронина, открывшего нарушение комбинированной симметрии (четности) в физике частиц, Шарля Рише, обнаружившего явление анафилаксии (аллергии), Александра Флеминга, открывшего антибактериальные свойства плесени, из которой впоследствии удалось выделить действующее вещество – пенициллин. Роль случая в своих открытиях отмечали также Кристиан де Дюв, обнаруживший лизосомы, Барух Бламберг, изучавший свойства австралийского антигена, Стенли Коэн, открывший фактор роста нервов (ФРН) в змеином яде. Не скрывали неожиданности (незапланированности) полученных результатов Эрнест Резерфорд, выяснивший существование атомного ядра, Николай Семенов, построивший теорию разветвленных химических реакций, Томас Чех, описавший каталитические свойства молекулы РНК.

Специалисты заметили, что случайные научные открытия часто могут выступать в роли фактора, инициирующего смену научных парадигм, то есть опровергать старую парадигму, освобождая место для появления новых теорий. К числу таких случайных прорывов можно отнести открытие связи между электричеством и магнетизмом, сделанное Христианом Эрстедом в 1820. Как известно, во время одного из экспериментов Х.Эрстед случайно заметил, что при включении электрического тока, бегущего по металлическому проводнику, стрелка магнитного компаса, расположенного рядом с проводником, начинает отклоняться от первоначального положения. Этот эксперимент ставил под сомнение прежде господствовавшую в физике парадигму. А.С.Майданов в книге «Искусство открытия» [10] пишет: «Господствовавшая тогда ньютоновская теория допускала только силы, которые действуют между материальными телами по соединяющей их

прямой. Открытие Эрстеда обнаружило силу иного рода: сила, действующая между магнитной стрелкой и проводником с током, оказалась направленной не по соединяющей их прямой, а перпендикулярно к ней. Теория Ньютона не могла предсказать такой результат и подсказать способ его получения. Напротив, она толкала к неверным приемам поиска связи электричества и магнетизма» [10].

Далее автор указывает: «Открытия принципиально новых явлений, по-видимому, всегда в какой-то мере случайны. Ведь даже необходимость имеет в качестве одной из форм своего проявления случайность. Возможности целенаправленного, а потому и ограниченного в отношении сферы и средств деятельности, исследования расширяются благодаря вторжению в это исследование случайных факторов. Детерминированная прошлым опытом поисковая деятельность как раз и может **выходить за рамки** этой деятельности благодаря, в частности, случайным обстоятельствам. Вполне очевидно, что случай не раз выручал исследователей, помогая им получить результат там, где оказывался бессильным сознательный поиск в рамках существующих представлений» [10].

Перечислим ряд известных концепций (парадигм), которые утратили свое значение и покинули «пьедестал науки» после того, как было сделано случайное открытие, показавшее ошибочность господствующей концепции.

- теория о том, что «природа боится пустоты», которую можно назвать парадигмой Аристотеля (опровергнута флорентийскими водопроводчиками, случайно заметившими, что насосы не могут поднять воду выше 18 локтей);

- теория А.Лавуазье о том, что все кислоты содержат кислород (опровергнута Жаном-Батистом Дюма, который случайно установил, что хлор способен вытеснить водород из химических соединений);

- теория Я.Берцелиуса о существовании «жизненной силы», о том, что в лабораторных условиях нельзя синтезировать сложные органические

соединения (опровергнута Фридрихом Велером, который случайно синтезировал мочевины из цианата аммония);

- теория Л.Пастера о том, что процесс ферментации невозможен без микробов (опровергнута Эдуардом Бухнером, который случайно открыл бесклеточное брожение);

- теория о том, что инертные элементы – гелий, неон, аргон и так далее – не могут взаимодействовать с другими химическими элементами (опровергнута Нилом Бартлеттом, который случайно обнаружил, что инертные газы могут окисляться гексафторидом платины);

- теория электромагнитных явлений, предложенная Г.Гельмгольцем (опровергнута Генрихом Герцем, который случайно обнаружил электромагнитные волны, предсказанные теорией Джеймса Максвелла);

- теория Вильяма Томсона, лорда Рэлея, о стабильности атома, о том, что атом не может распадаться на части (опровергнута Анри Беккерелем, случайно открывшим радиоактивность);

- теория Эрнеста Резерфорда о невозможности практического использования ядерной энергии (опровергнута Отто Ханом, который случайно открыл деление атома урана);

- теория стационарной Вселенной, предложенная Альбертом Эйнштейном и развитая Фредом Хойлом (опровергнута Арно Пензиасом и Робертом Вильсоном, которые случайно открыли космическое реликтовое излучение);

- теория Артура Эддингтона о том, что гравитационные волны не могут существовать в нашем мире (ее опровергли Рассел Халс и Джозеф Тейлор, которые случайно открыли двойной пульсар PSR 1913+16, позволивший косвенно подтвердить реальность гравитационных волн);

- теория комбинированной симметрии в физике частиц, предложенная Львом Ландау (опровергнута Джеймсом Кронином, который случайно нашел условия нарушения комбинированной симметрии).

Анализ роли случайных открытий в изменении наших теорий, в переходе от одной парадигмы к другой, заставил многих историков науки серьезно отнестись к фактору случая в научном поиске. В.Азерников в статье «Продолжение следует» [11] подчеркивает: «Случайных находок в науке довольно много. Более того, можно утверждать, что большинство экспериментальных открытий сделано именно случайно, потому что никаким другим способом их сделать бы не удалось. <...> Случайная находка не является компрометирующим обстоятельством, она входит в технологию научной работы и так же правомерна, как и исполнение любого пункта годового плана. Кстати, в каждом плане всё же остается место для узаконивания случайных находок, иначе наука застряла бы на месте. Вероятно, каждый настоящий экспериментатор, начиная свой рабочий день, в глубине души ждет ее, эту самую случайность, после которой всё либо переворачивается с ног на голову, либо, наоборот, утверждается в существующей позиции – в зависимости от того, в каком положении находилась работа до этого» [11, с.68].

Аналогичную мысль высказывает Ю.В.Ходаков в книге «Как рождаются научные открытия» [12]: «...Если из истории точных наук вычеркнуть случайные открытия, она приобретет мистический характер, перестанет существовать, рассыпится, как карточный домик» [12].

3. Типичные ситуации, в которых совершаются случайные открытия

На основе анализа самых разных научных открытий, которые по своей природе относятся к категории случайных (незапланированных), можно выделить, по меньшей мере, десять типичных ситуаций, в которых осуществляются эти открытия. Ниже приводится описание этих ситуаций, позволяющих определить ключевой механизм (основную причину) неожиданных находок.

Ученый забывает выключить экспериментальное оборудование.

Вильгельм Рентген открыл лучи, названные его именем, благодаря следующему обстоятельству. Завершив исследование, проводимое в течение дня и покидая свой рабочий кабинет, он забыл выключить разрядную (катодно-лучевую) трубку. Погасив свет в кабинете, ученый заметил свечение фосфоресцирующего экрана, расположенного вблизи разрядной трубки. Этот экран предназначался для изучения других лучей – ультрафиолетовых. Экран засветился, хотя разрядная трубка была закрыта непрозрачным черным картоном. Встал вопрос: почему разрядная трубка, закрытая картоном, вызывает свечение экрана? Исследование этого вопроса позволило установить, что разрядная трубка испускает, помимо катодных лучей, еще и невидимые лучи, для которых непрозрачный черный картон не является препятствием. Эти невидимые лучи и были впоследствии названы рентгеновскими. Таким образом, ключевым фактором открытия явилось то, что В.Рентген, покидая кабинет, забыл отключить свою разрядную трубку. Хотя, конечно, существенное значение имело и то, что рядом с указанной трубкой оказался фосфоресцирующий экран, предназначенный для исследования ультрафиолетовых лучей.

Ученый пытается проверить неверную (ошибочную) гипотезу. Анри Беккерель открыл радиоактивность благодаря тому, что поставил перед собой задачу проверить неверную гипотезу. Конечно, до проведения экспериментов нельзя было сказать, что проверяемая гипотеза является ошибочной. Автором этой гипотезы был великий французский математик и механик Анри Пуанкаре, который предположил, что источником рентгеновских лучей могут быть любые фосфоресцирующие вещества. В ходе эксперимента А.Беккерель взял соль урана, поместил ее в фольгу и намеревался подвергнуть материал воздействию солнечного света. Однако пасмурная погода помешала ему осуществить задуманное, и он убрал соль урана в шкаф. Когда же спустя время ученый достал материал, чтобы продолжить опыт, то обнаружил, что

указанная урановая соль оставила изображение на фотопластинке. Это свидетельствовало о том, что урановая соль испускает какое-то излучение в темноте, без воздействия солнечного света. В результате был сделан важный шаг к открытию явления радиоактивности (явления распада атомов, во что не верил Вильям Томсон, лорд Кельвин).

Ученый не спешит прервать опыт, который не дал сиюминутных результатов. Часто мы полагаем, что эксперимент (продолжительность которого мы намеренно ограничиваем) должен сразу дать ответ на тот или иной вопрос. Если мы не получаем значимой информации в течение короткого времени, мы склонны прервать эксперимент и приступить к новым исследованиям. Однако в ряде случаев полезно отказаться от такой торопливости (вы можете, не прерывая опыт, немного отвлечься, например, сходить в буфет и выпить кофе). Историкам науки известно, что в 1902 г. именно отказ от подобной «торопливости» позволил российскому физиологу Алексею Александровичу Кулябко (1866-1930) впервые в мире оживить сердце человека через 20 часов после его смерти.

В.В.Александрин в статье «Падре Реанимационе» [13] пишет: «Неисповедимы пути научных идей. В 1905 году русскому физиологу А.Кулябко доставили сердца детей, умерших от пневмонии в одной из петербургских клиник. Кулябко подвесил их на трапеции и подвел к каждому теплый физиологический раствор, насыщенный кислородом. Жидкость попадала по остатку аорты в сердечные сосуды, проходила через капиллярную сеть и стекала через вены. Первые опыты ученого ни к чему не привели: десятиминутная прокачка раствора не оживляла мертвые сердца. Тогда, «по законам жанра», профессор вышел в буфет попить чайку, забыв снять сердце с трапеции. Вернувшись через полчаса, он застал орган сокращающимся – через сутки после смерти» [13, с.28].

Ученый поручает ассистенту выполнить простой опыт. Когда Николай Николаевич Семенов поручал своей сотруднице (Зинаиде Вальта)

провести простой эксперимент – выяснить, как изменяется свечение фосфора при изменении давления кислорода, - он не рассматривал этот опыт как многообещающий. Более того, Н.Н.Семенов не считал его интересным, поскольку он не был связан с основной тематикой исследований лаборатории, которую возглавлял ученый. Можно сказать, используя шахматную терминологию, что опыт, порученный З.Вальта, не представлял собой фигуру, на которую бы делалась ставка. Однако именно изучение взаимодействия фосфора с кислородом привело к открытию разветвленных цепных химических реакций - открытию, которое принесло Николаю Николаевичу Нобелевскую премию по химии за 1956 г.

Н.Н.Семенов в статье «Таким образом, я пришел к идее...» [14] вспоминает об эксперименте, выполнение которого было поручено Зинаиде Вальта: «Тема эта не являлась развитием других наших работ и идей. Она была выбрана случайно. И, признаться, не очень меня интересовала. Если бы я знал, что двойная случайность – принятие в аспирантуру Вальта и поручение ей именно этой темы – определит в дальнейшем в значительной мере работу всего нашего коллектива! Конечно, разветвленные цепные реакции все равно были бы неизбежно открыты в скором времени, но то, что именно мы оказались пионерами этой важнейшей области химии и физики, явилось делом случая» [14, с.40].

Ученый не смог исключить воздействие посторонних факторов. Как правило, при постановке эксперимента исследователь старается сделать так, чтобы никакие посторонние факторы не влияли на исследование и не искажали его результаты. Но иногда воздействие этих факторов становится преимуществом, а не недостатком (дефектом). Александр Флеминг открыл антибактериальное действие плесени благодаря тому, что однажды в открытое окно его лаборатории проникли споры грибковой плесени. В это время он изучал особенности роста и размножения патогенных микробов (стафилококков). Плесень, попавшая в чашку с этими микробами, стала

подавлять их рост и размножение. Позже исследователям (которые пошли по следам А.Флеминга) удалось выделить из плесени действующее начало – пенициллин.

В.В.Мазин и Л.С.Шашкова в книге «Грибы, растения и люди» [15] приводят слова А.Флеминга о сделанном им открытии: «Со времен Пастера и Листера ученые пытались убить один микроб другим. Идея была правильная, но для ее осуществления пришлось ждать дня, когда **фортуна решила**, что споры плесени заразят одну из исследуемых мной культур, а потом несколько лет спустя настал и другой день, когда химики занялись веществом, выделяемым этой плесенью, и дали нам чистый пенициллин» [15, с.192].

Ученый ставит эксперимент, не имея конкретной цели (по принципу «надо попробовать»). Эту ситуацию, в которой также могут совершаться случайные открытия, описывали многие историки науки. Например, А.К.Сухотин в книге «Превратности научных идей» [16] констатирует: «Нередко исследователь, проводя опыт, вначале никакой, хотя бы отдаленной, научной цели не ставит, а ставит эксперимент, можно сказать, бездумно: что получится, то и получится. Это неопределенно-значный эксперимент с большим разбросом предполагаемых исходов, - настоящее чудачество. <...> В самом начале текущего века русский ботаник М.Цвет, взяв однажды раствор хлорофилла, пропустил его через стеклянную колонку, набитую мелом. Зачем это делал, и сам не знал, пропустил без всякого умысла, как бы забавы ради. К своему удивлению, он обнаружил, что зеленый цвет хлорофилла разделился на два оттенка – зеленый и желтый. Значение полученного результата осознали не сразу» [16, с.124].

Далее автор отмечает: «Между тем опыт принес новый метод, который стал впоследствии широко использоваться в экспериментальной химии. Порой он просто незаменим. В частности, в ситуации, когда исследователь не видит цвет раствора (реакция идет в закрытом от глаз сосуде). Однако с помощью метода можно определить цвет, зная лишь участвующие во

взаимодействии компоненты» [16, с.125]. «В известном смысле продолжением исследований М.Цвета является разработка английскими биохимиками Л.Мартином и Р.Синджем эффективного способа разделения сложных химических смесей – метода распределительной хроматографии, удостоенного в 1952 году Нобелевской премии» [16, с.125].

Ученый ставит контрольный опыт, используя средства, которые приводят к случайному открытию. Примером такой ситуации является случайная находка американского биолога Маршалла Ниренберга. В 1968 г. он получил Нобелевскую премию за фундаментальное достижение – расшифровку генетического кода. Но как он его расшифровал? Какое открытие привело его к этому успеху? Как ни удивительно, успех М.Ниренберга был обусловлен случайным открытием, которое стало возможно благодаря постановке контрольного опыта. М.Ниренберг мог благословить тот день, когда ученые изобрели такую вещь, как контрольный опыт! Ключевая находка, сделанная М.Ниренбергом, заключалась в том, что он совершенно случайно обнаружил следующий факт: полинуклеотид под названием «полиурацил» инициирует (запускает) синтез белка под названием «полифенилаланин».

Сначала М.Ниренберг взял РНК вируса табачной мозаики (ВТМ), поместил ее в бесклеточную систему синтеза белка и обнаружил, что синтез белка идет: РНК ВТМ стимулировала включение радиоактивных аминокислот в белок. Этот положительный опыт М.Ниренберга служил подтверждением гипотезы других ученых об участии информационной РНК в качестве посредника в синтезе белков. После этого М.Ниренберг решил поставить **контрольный опыт**, желая взять какой-нибудь полинуклеотид и в финале опыта обнаружить его неспособность обеспечивать производство белка. Когда ученый выбрал в качестве полинуклеотида полиурацил и бросил его в реакционную смесь для синтеза белков, он ожидал, что этого синтеза не будет, поскольку считал данный полинуклеотид некодирующим, то есть не

имеющим никакого отношения к кодированию синтеза аминокислот. Однако, вопреки ожиданию Ниренберга, синтез пошел, был синтезирован белок полифенилаланин. Ученый понял, что ему удалось обнаружить первый кодон, запускающий синтез конкретного белка. Одновременно он понял, что если последовательно помещать в реакционную среду различные полинуклеотиды и наблюдать, какие белки при этом синтезируются, то можно определить все кодоны, то есть полностью расшифровать генетический код! Вот почему сразу после своего случайного открытия М.Ниренберг поспешил на Международный биохимический конгресс, проходивший в 1961 г. в Москве (нужно было срочно сообщить о полученном результате).

И.Харгиттай в книге «Откровенная наука» [17] приводит слова Шарля Вайсмана о роли контрольного опыта в открытии М.Ниренберга: «Ниренберг хотел использовать РНК вируса табачной мозаики (ВТМ) в качестве информационной РНК для синтеза белка оболочки ВТМ. Он достал у Хайнца Френкель-Конрата ВТМ, поместил ее в бесклеточную систему синтеза белка, которую сам разработал, и обнаружил, что она стимулирует включение радиоактивных аминокислот в белок, как и ожидалось согласно гипотезе об информационной РНК. Но потом он и Маттэи провели эксперимент, который поначалу считался всего лишь **контрольным**. Они должны были взять РНК или какой-нибудь другой полинуклеотид, который, как ожидалось, ничего не кодирует, и показать, что он не стимулирует синтез белка. Они выбрали poly(U)» [17, с.434]. Что произошло дальше, читатель уже знает.

Еще одно непреднамеренное открытие, сделанное благодаря проведению контрольного опыта, - открытие РНК-интерференции. Авторами открытия являются американские ученые Эндрю Файер и Крэйг Мелло, которые в 2006 г. удостоены Нобелевской премии по физиологии и медицине. Э.Файер и К.Мелло в своих экспериментах вводили в организм червей-нематод *Caenorhabditis elegans* одноцепочечную РНК, зная, что такая РНК должна подавлять экспрессию генов червей. Каково же было их удивление, когда

выяснилось, что двухцепочечные РНК, вводимые (в качестве контроля) в клетки тех же червей, также способны блокировать активность генов! Так была обнаружена РНК-интерференция, нашедшая массу применений в биологии. Читатель найдет описание этого случайного открытия в статье Ю.Н.Елдышева «Нобелевские премии - 2006» [18].

Ученый ставит опыт, в котором происходит авария. Казалось бы, авария, происходящая во время проведения эксперимента, не может иметь положительных последствий с точки зрения результатов научного поиска. Любая авария приводит к поломке приборов, выходу из строя ценного оборудования, заставляет исследователей остановить эксперимент. Но бывают исключения из этого правила, и этим исключением в свое время воспользовался Клинтон Дэвиссон, открывший дифракцию электронов и получивший в 1937 г. за этот исследовательский успех Нобелевскую премию по физике.

И.С.Дмитриев в книге «Электрон глазами химика» [19] пишет: «Когда Дэвиссон в 1927 г., на этот раз совместно с Джермером, проводил опыт по рассеянию электронов от никелевой пластинки, в установку случайно попал воздух и поверхность металла окислилась. Пришлось удалять оксидную пленку отжигом кристалла в высокотемпературной печи в восстановительной среде, после чего опыт продолжили. Однако результаты его стали иными. Вместо монотонного (или почти монотонного) изменения интенсивности рассеянных электронов от угла наблюдались ярко выраженные максимумы и минимумы, положение которых зависело от энергии электронов. <...> Причина столь резкого изменения картины рассеяния **после аварии** состояла в образовании в результате отжига монокристаллов никеля, которые служили своего рода дифракционными решетками» [19, с.21].

Об этом же сообщает А.С.Майданов в книге «Искусство открытия» [10]: «Фактором, способствовавшим экспериментальному подтверждению волновой природы электрона, была **авария**, происшедшая во время опыта

К.Д.Дэвиссона по рассеянию электронов. Эта авария помогла получить мишень, которая и явилась ключевым элементом открытия» [10].

Ученый извлекает подсказку из случайного разговора. Случайный разговор помог датскому физика Нильсу Бору (Нобелевская премия, 1922 г.) разработать квантовую модель атома. Ключевой момент этой теории – описание условий, при которых атом излучает кванты света. Нильс Бор установил, что электрон, вращающийся вокруг атомного ядра, излучает квант света лишь тогда, когда «перескакивает» с одной орбиты на другую. Как же выдающийся датский физик установил это? В этом ему помогло изучение спектроскопии, а именно спектральных формул Бальмера, Ридберга и Ритца. А как Н.Бор догадался, что именно в спектроскопии нужно искать ответы, которые позволят разработать квантовую модель атома? На верный путь его навел случайный разговор с однокашником по университету Хансом Хансеном.

А.Б.Мигдал в статье «Нильс Бор и квантовая физика» [20] повествует: «Событие, которое стало для Бора последним толчком, произошло в начале февраля 1913 г. По чистой случайности он встретил своего приятеля студенческих лет Ханса Хансена, специалиста по спектроскопии. Когда Бор рассказывал ему свои идеи строения вещества на основе планетарного атома с устойчивыми по неведомым причинам орбитами, Хансен спросил: «А как твоя теория объясняет спектральные формулы?» И тут, к его огромному удивлению, обнаружилось, что Бор ничего не знает о спектральных формулах, полученных Бальмером (1885), Ридбергом (1890) и Ритцем (1908). Бор вспоминал: «Как только я увидел формулу Бальмера, мне все сразу же стало ясно». Это было вспышкой, осветившей всю картину, - менее чем за месяц была готова первая, самая существенная часть работы «О строении атомов и молекул» [20, с.318-319].

Аналогичные сведения можно найти в книге Д.Данина «Вероятностный мир» [21], где случайный разговор Н.Бора с однокашником Х.Хансеном описывается на стр.76-77.

Нечто аналогичное произошло с автором этих строк, когда мы занимались анализом индуктивных открытий (открытий, в которых ключевая идея возникает на основе индуктивного обобщения экспериментальных данных). Правда, решающую роль сыграл не случайный разговор, а неожиданная встреча с книгой! В 2012 г. мы приступили к изучению индуктивных открытий в области математики. На каком-то этапе работы мы задались целью рассмотреть математические открытия американского математика Стивена Смейла (лауреата премии Филдса за 1966 г.). Совершенно случайно мы натолкнулись на книгу «Современные проблемы хаоса и нелинейности» [22], в которой содержался список нерешенных математических проблем, составленный С.Смейлом по просьбе В.И.Арнольда. Когда мы ознакомились с 18-й проблемой С.Смейла из этого списка, а именно с вопросом: каковы пределы интеллекта (как искусственного, так и человека)? – молниеносно появилась мысль, что одним из этих пределов является индукция. Та индукция, вероятностный характер которой был известен еще Г.Лейбницу (и не только ему). А далее стало ясно, что, наряду с индукцией, другим фактором, препятствующим формализации научного творчества, являются случайные открытия. Но вернемся к рассмотрению типичных ситуаций, в которых возникают непреднамеренные открытия, и опишем еще одну ситуацию.

Ученый совершает экспериментальную ошибку. Выше мы указали, что в список Нобелевских лауреатов, чьи открытия были сделаны благодаря фактору случая, должны войти Алан Хигер, Алан Мак-Диармид и Хидеки Ширакава (Сиракава), награжденные Нобелевской премией по химии в 2000 году. Этим исследователям посчастливилось изобрести электропроводящие полимеры, то есть открыть условия, при которых обычный полимер (не способный проводить электрический ток) становится электропроводящим.

Как же ученые открыли эти условия? В результате экспериментальной ошибки, допущенной одним из сотрудников (ассистентов) Хидеки Ширакавы.

В.А.Марихин в статье «Синтетические металлы» [23] пишет: «А в 1971 году профессор Токийского технологического института Хидеки Ширакава дал своему аспиранту задание синтезировать полимер ацетилена. Впервые полиацетилен был получен еще в 1955 году в виде темного порошка, не обладающего никакими особо выдающимися свойствами. Однако аспирант по ошибке добавил в реакционную смесь в 1000 раз больше катализатора, чем требовалось по методике (наверное, перепутал граммы с миллиграммами), в результате чего вместо темного порошка получил роскошную пленку с металлическим блеском. Едва взглянув на эту пленку, Ширакава подумал, что она может послужить основой для создания полимеров, обладающих свойствами металлических проводников» [23, с.11].

Аналогичные сведения об истории открытия можно найти в статье Ричарда Канера и Алана Мак-Диармида «Электропроводящие полимеры» [24], где авторы сообщают: «Открытие электропроводящих полимеров произошло случайно. В начале 70-х годов аспирант лаборатории Хидеки Сиракавы при Токийском технологическом институте пытался получить полимер под названием «полиацетилен» из обычного ацетилена – горючего газа, используемого при сварке. Этот полимер в виде порошка темного цвета впервые был получен в 1955 г., но о нем мало что было известно. Вместо темного порошка аспирант из лаборатории Сиракавы получил блестящую серебристую пленку, похожую на алюминиевую фольгу, но эластичную. Просматривая свои записи, аспирант заметил допущенную ошибку: оказалось, что он добавил в 1000 раз больше катализатора, чем требовалось по методике. То, что он получил, было, конечно, полиацетиленом, но по форме отличным от полиацетилена, известного ранее» [24, с.50].

Итак, мы рассмотрели 10 типичных (часто встречающихся) ситуаций, в которых совершаются «серендипные», т.е. незапланированные научные

открытия. Возникает вопрос: можно ли наделить искусственный интеллект способностью делать подобные открытия? Можно ли передать вычислительной машине функцию замечать находки, которые не предусматривались задачами и целями основной программы исследований? Иначе говоря, можно ли создать компьютер, не пренебрегающий принципом серендипити («серендипный» компьютер)?

4. Как научить искусственный интеллект делать случайные открытия?

Основная черта непреднамеренных открытий – обнаружение того, что не предусматривалось программой исследований, не имело непосредственной связи с задачами и целями данной программы. Но, как ни удивительно, путь к неожиданным находкам пролегает через реализацию этих задач и целей. Следовательно, чтобы сделать «серендипное» открытие, нужно работать в рамках намеченной программы, одновременно учитывая возможность обнаружения того, что не вписывается в нее. Отсюда следует, что искусственный интеллект должен научиться планировать предстоящее исследование, составлять программу работы над какой-то проблемой. Иначе говоря, машина должна научиться формулировать цели и задачи экспериментальной и иной научной деятельности.

Является ли данное требование (условие) обязательным? Нет, не является. Вполне можно допустить, что машина не составляла программу работы, что формулировкой целей и задач занимался человек. А машина была всего лишь одним из участников проводимого исследования, подобно тому, как в любом научном коллективе есть те, кто планирует серию необходимых опытов, то есть составляет программу, и те, кто непосредственно ставит эти опыты. Вот они-то – сотрудники, ставящие запланированные эксперименты, и сталкиваются первыми с явлениями, о которых не думал и не догадывался никто, в том числе руководитель исследования (научного проекта).

Голландский ученый Хейке Камерлинг-Оннес получил в 1913 г. Нобелевскую премию по физике за открытие явления сверхпроводимости, но впервые столкнулся с этим явлением не он, а его сотрудник Г.Холст. Он был первым, кто наблюдал эффект исчезновения электрического сопротивления у ртути вблизи абсолютного температурного нуля. Роль Г.Холста, который в дальнейшем стал профессором Лейденского университета, упоминается в статье В.Л.Гинзбурга «Несколько замечаний об изучении сверхпроводимости» [25, с.187].

Американский биолог Томас Хант Морган получил в 1933 г. Нобелевскую премию по физиологии и медицине за создание хромосомной теории наследственности, но убедительное доказательство этой теории получил в 1914-1916 гг. его сотрудник Кальвин Бриджес, когда открыл явление нерасхождения хромосом. А первым, кто на практике применил теоретический вывод о линейном расположении генов в хромосоме, был еще один сотрудник Т.Моргана, Альфред Стертевант [26].

Немецкий ученый Ханс Шпеман получил в 1935 г. Нобелевскую премию по физиологии и медицине за открытие эмбрионального организатора и разработку концепции эмбриональной индукции, но решающий эксперимент, позволивший установить существование эмбрионального организатора, поставил не он, а его сотрудница Хильда Мангольд. Роль Х.Мангольд в постановке этого эксперимента обсуждается в статье Ю.Н.Городилова «Организатор Шпемана: его источники и его производные» [27].

Американский микробиолог Зельман Ваксман получил в 1952 г. Нобелевскую премию по физиологии и медицине за открытие стрептомицина – антибиотика, эффективного против туберкулеза, но впервые выделил этот антибиотик и исследовал его противотуберкулезное действие не он, а его аспирант Альберт Шац (Albert Schatz, 1920-2005). Об этом, в частности, сообщается в книге А.А.Капто «Анналы урологии» [28, с.286].

Испано-американский исследователь Северо Очоа получил в 1959 г. Нобелевскую премию за открытие полинуклеотидной фосфорилазы, но впервые этот фермент, участвующий в распаде РНК, выделила его аспирантка Марианна Грюнберг-Манаго (1921-2013). Впоследствии она станет президентом Международного биохимического союза (1982-1988). Роль М.Грюнберг-Манаго в открытии, которое принесло Нобелевскую премию ее руководителю, описывается во многих работах, в том числе в [29].

Американский ученый Маршалл Ниренберг получил в 1968 г. Нобелевскую премию за расшифровку генетического кода, но первая находка, позволившая достичь этой цели, была сделана его сотрудником Генрихом Маттэи. Он обнаружил, что полиурацил запускает синтез белка полифенилаланина. Как сообщает И.Харгиттай в книге «Откровенная наука» [17], «именно Маттэи первым провел эксперимент с полиурацилом. После этого эксперимента он тут же позвонил Ниренбергу, и тот спешно вернулся в Бетесду» [17, с.129].

Британский астроном Энтони Хьюиш получил в 1974 г. Нобелевскую премию по физике за открытие нейтронных звезд (радиопульсаров), но первые сигналы этих пульсаров получила его аспирантка Джоселин Белл (род. 1943 г.). Фред Хойл, известный британский астрофизик, недоумевал по поводу того, почему Нобелевский комитет не вручил ей премию совместно с Э.Хьюишем.

Таким образом, вполне можно представить ситуацию, при которой искусственный интеллект не составляет никаких программ исследования, не определяет цели и задачи предстоящих экспериментов, а всего лишь участвует в деятельности научной лаборатории. Что же необходимо, чтобы машина могла замечать сюрпризы (неожиданные события), которые порой «подбрасывает» тот или иной эксперимент?

Во-первых, необходимо, чтобы искусственный интеллект обладал широкой эрудицией. Эта эрудиция должна включать в себя информацию (по крайней мере, в той области знаний, в которой проводятся эксперименты) о

том, что известно современной науке и что остается непознанным, неисследованным в ней. Если вы осведомлены о достижениях своей области знаний, то при неожиданном столкновении с каким-то новым фактом вы без труда определите, что перед вами действительно новый факт. Для этого достаточно сопоставить (сравнить) имеющиеся у вас знания с новым явлением, которое случайно обнаружено в процессе исследований. Следовательно, машина должна постоянно сопоставлять информацию, получаемую в ходе экспериментов, с теми данными, которые хранятся в ее памяти. При отсутствии такого сравнения новый факт останется незамеченным (не удастся понять, что он новый).

Во-вторых, необходимо, чтобы искусственный интеллект умел сосредотачиваться на изучении случайных находок, концентрировать усилия на анализе аномальных (не вписывающихся в избранную программу) явлений. Иногда это заставляет резко менять направление поисков, приостанавливать избранную программу и переключаться на решение других проблем. Такое переключение, как правило, требует не только времени, но и дополнительных финансовых затрат, поскольку любой новый эксперимент имеет определенную денежную стоимость. Важно не отмахиваться от неожиданных явлений, попавших в поле зрения, не игнорировать то, что еще не поддается объяснению (и, помимо этого, избегать чрезмерной приверженности старым концепциям, мешающим увидеть новое).

История науки знает множество случаев, когда ученые проходили мимо случайных находок, не придавая им значения в силу тех или иных причин. Например, в 1930-х гг. американские физики из Беркли могли опередить супругов Жолио-Кюри в открытии искусственной радиоактивности, но проигнорировали показания счетчиков, которые свидетельствовали об этой радиоактивности. Анатолий Абрагам в книге «Время вспять...» [30] пишет: «В тридцатых годах физики-ядерщики в Беркли избавились от паразитного раздражающего треска своих счетчиков с помощью тщательного

экранирования, только чтобы убедиться, прочитав публикацию супругов Жолио, что они закрыли глаза (вернее, уши) и прозевали открытие искусственной радиоактивности» [30, с.200].

Правда, и сами супруги Жолио-Кюри упустили возможность сделать, по меньшей мере, три открытия Нобелевского уровня. А.Абрагам в той же книге «Время вспять...» [30] сообщает: «Хотя главные работы Фредерика и Ирины Жолио – открытие искусственной радиоактивности, вознагражденное Нобелевской премией (и в меньшей степени работы тройки – Жолио, Гальбана и Коварски), - являются частью научной истории страны, мало кто знает, что супруги Жолио прошли мимо трех великих открытий, тоже вознагражденных Нобелевской премией. Прежде всего – открытие нейтрона, буквально оброненное супругами Жолио в корзину англичанина Чэдвика из-за их неспособности правильно проанализировать результаты своих опытов. Затем открытие позитрона, траектории которых Жолио наблюдал в своей камере Вильсона, но истолковал как следы электронов, движущихся к источнику. <...> Наконец, Ирина Жолио наблюдала, что химические свойства некоторых предполагаемых трансурановых элементов, полученных нейтронным облучением урана, поразительно похожи на свойства лантана; но, как и другие крупные радиохимики того времени, она не смогла сделать последнего решающего шага и признать, что то, что так смахивало на лантан, и есть лантан, и тем самым прошла мимо открытия ядерного деления, наложившего неизгладимую печать на двадцатое столетие» [30, с.157].

В-третьих, необходимо, чтобы искусственный интеллект умел обобщать информацию, предоставляемую случайной находкой, выявлять причинно-следственную связь между неожиданным экспериментальным результатом и теми обстоятельствами, которые обусловили этот результат. Если опыт позволил случайно обнаружить какое-то явление, нужно убедиться, что перед нами не артефакт, а реальный физический (или биологический) эффект. Для этого требуется неоднократно воспроизвести (повторить) поставленный опыт,

разобраться во всех его деталях, вскрыть механизмы, действие которых определяет (детерминирует) новое явление. Повторение опыта позволяет, в том числе, проверить сформулированную гипотезу о наличии причинно-следственной связи между новым явлением и теми факторами, которые ему предшествуют. Правильному исследованию способствует использование принципов индуктивного анализа фактов, которые впервые изложены британским ученым Джоном Стюартом Миллем (1806-1873) в знаменитой книге «Система логики силлогистической и индуктивной», изданной в 1843 г. Д.С.Милль выделил пять правил индуктивного (опытного) исследования: 1) метод сходства; 2) метод различия; 3) соединенный метод сходства и различия; 4) метод остатков; 5) метод сопутствующих изменений. Подробное описание этих правил можно найти в книге В.И.Курбатова «Логика» [31, с.201-202].

Известно, что немецкий химик Карл Циглер открыл катализаторы, обеспечивающие синтез полиэтилена, благодаря случайной находке, сделанной лаборантом. В 1963 г. этот исследовательский успех принес К.Циглеру Нобелевскую премию по химии. Г.Б.Шульпин в книге «Мир необычных молекул» [32] пишет: «Катализаторы полимеризации этилена были открыты в начале 50-х годов, и притом **совершенно случайно**. Сотрудники западногерманского ученого К.Циглера проводили реакции этилена с алюминийорганическими соединениями. Однажды лаборант не удосужился помыть как следует автоклав после каких-то опытов с соединениями никеля, загрузил в него алюминийорганическое соединение, этилен и... обнаружил, что весь этилен легко и быстро заполимеризовался. Окрыленные успехом исследователи начали перебирать все возможные соединения переходных металлов, которые добавляли в качестве второго компонента к триалкилалюминию. Вскоре нашли, что наиболее эффективны хлоридные комплексы титана» [32, с.142].

Аналогичное описание истории открытия К.Циглера содержится в книге В.А.Афанасьева и Г.Е.Заикова «В мире катализа» [33]. А.Л.Бучаченко в

предисловии к данной книге указывает: «В прошлом, да и в недалеком прошлом, многие высокоэффективные катализаторы находили **случайно** и совсем не там, где искали и не тот, кто искал. Совершенно неожиданно были, например, открыты катализаторы стереоспецифической полимеризации Циглера. Их открытие связано с тем, что в недостаточно хорошо отмытом реакторе остались продукты предшествующих процессов. В результате этой «неаккуратности» и глубокого внимания химиков-исследователей, которые не прошли мимо неожиданно полученного эффекта, произошло замечательное событие – были найдены катализаторы, открывшие новую важную страницу в науке и практике производства полимеров» [33, с.3].

Итак, разобьем открытие Карла Циглера на отдельные стадии. Сначала лаборант плохо помыл автоклав, из-за чего в нем остались какие-то соединения никеля. Не зная об этом, К.Циглер проводит очередной эксперимент и обнаруживает полимеризацию этилена в автоклаве. Что делает будущий Нобелевский лауреат? Он начинает искать причинно-следственную связь, то есть выяснять, по какой причине произошла полимеризация этилена. Его подозрения падают на соединения никеля, оставшиеся в автоклаве после предыдущего опыта. Чтобы проверить свою гипотезу (свои подозрения), К.Циглер воспроизводит эксперимент, помещая в камеру этилен и соединения никеля. Гипотеза подтверждается, после чего ученый делает индуктивный вывод о том, что соединения никеля могут выступать в роли катализатора полимеризации этилена. А далее немецкий химик использует мыслительную операцию, называемую «аналогией»: он делает заключение, что по аналогии с соединениями никеля должны существовать другие соединения металлов, способные эффективно запускать процесс полимеризации этилена. К.Циглер ставит новую и весьма масштабную серию экспериментов и находит, что хлоридные комплексы титана еще лучше полимеризуют этилен, чем соединения никеля. Вот такой работе должен обучиться искусственный

интеллект, чтобы делать случайные открытия и извлекать максимальную пользу из этих находок!

Если машина (роботизированная система, наделенная искусственным разумом) научится делать подобные открытия, то типичные ситуации «серендипных» находок, рассмотренные нами применительно к человеку, вполне могут стать типичными ситуациями таких же находок, сделанных компьютером. И тогда не придется удивляться попыткам систематизации (классификации) механизмов возникновения случайных исследовательских успехов, принадлежащих искусственному интеллекту. Или встречать в работах, посвященных анализу этих механизмов, таблицы, эквивалентные по своему содержанию той, которую мы приводим ниже.

Таблица 1. Возможные условия (обстоятельства), в которых могли бы совершаться случайные открытия, принадлежащие искусственному интеллекту (ИИ)

№	Описание условия, в котором возможно случайное открытие
1.	ИИ забывает выключить экспериментальное оборудование
2.	ИИ пытается проверить неверную гипотезу
3.	ИИ не спешит прервать опыт, который не дал сиюминутных результатов
4.	ИИ берется выполнить простой опыт (с которым не связывает никаких надежд)
5.	ИИ не смог исключить воздействие посторонних факторов
6.	ИИ ставит эксперимент, не имея конкретной цели (по принципу «надо попробовать»)
7.	ИИ ставит контрольный опыт, используя средства, которые приводят к случайному открытию
8.	ИИ ставит опыт, в котором происходит авария
9.	ИИ извлекает подсказку из случайного разговора
10.	ИИ совершает экспериментальную ошибку

Примечание: ситуации, описанные в данной таблице, не исчерпывают список обстоятельств, сопутствующих случайному открытию. В реальности подобных ситуаций может быть чрезвычайно много (мы всего лишь хотели дать общее представление о механизмах возникновения эпизодов «serendipity» в науке).

5. Заключение

Человек может выбрать вариант эволюции науки, при котором искусственный интеллект обретет высокую степень самостоятельности (свободы). В этом случае машины станут такими же полноценными субъектами научных исследований, как и люди. Машины будут обладать полномочиями самостоятельно выбирать цели и задачи исследования, разрабатывать программу решения определенной научной проблемы, а также методы достижения искомого результата. Машины получат возможность эффективно пользоваться индуктивной логикой, публиковать статьи и книги о сделанных открытиях и заявлять о своем приоритете (первенстве) на тот или иной научный результат. При таких обстоятельствах нельзя исключать, что машины будут посещать научные конференции и обсуждать перспективы тех или иных теорий (концепций) с таким же энтузиазмом, как это делают ученые-люди. Кстати, мобильный робот уже присутствовал на конференции и выступал там с докладом. Стюарт Рассел и Питер Норвиг в книге «Искусственный интеллект» [34] пишут о соревновании мобильных роботов, проведенном в 2002 г.: «Наблюдаемый прогресс остается стабильным и весьма впечатляющим: в одном из последних соревнований (2002 год) перед роботами стояла задача войти в комплекс зданий, где проводилась конференция, найти путь к бюро регистрации, зарегистрироваться на конференции и произнести речь» [34, с.1238].

Однако, скорее всего, человек остановит свой выбор на другом сценарии развития научного знания. Будут приняты меры, в соответствии с которыми самостоятельность (свобода) интеллектуальных машин будет существенно ограничена. Компьютерам с элементами разумной деятельности не позволят выбирать цели и задачи исследования. Их роль будет сводиться к тому, чтобы помогать человеку в проведении научного поиска, участвовать в постановке некоторых экспериментов. Их не станут наделять всеми когнитивными

функциями, которыми обладает человеческий разум. Но функцией осознавать (понимать) новизну явления, обнаруженного в ходе поиска, скорее всего, наделят. Это одна из самых ценных функций. Она позволяет, продвигаясь шаг за шагом, от эксперимента к эксперименту, расширять границы тех знаний, которыми мы обладаем, находить практическое приложение (применение) этим знаниям, стимулировать прогресс множества технологий.

Но, даже если машина освоит «искусство случайных открытий», она будет делать их без возможности исчерпывающим образом формализовать (алгоритмизировать) это искусство. Как мы отметили в [8, 9], фактор случая, присутствующий в научном поиске, нельзя формализовать, поскольку подлинное научное творчество предполагает выход за границы известного, а в этой «неизведанной области» нельзя знать заранее, с какими аспектами (закономерностями) природы мы столкнемся. Изучая катодные лучи, образующиеся в разрядной трубке, мог ли Вильгельм Рентген заранее знать, что на этом пути ему посчастливится открыть рентгеновские волны? Случайные открытия – один из факторов, позволяющих ответить на вопрос американского математика Стивена Смейла о пределах интеллекта (пределах алгоритмизации интеллекта). Этот фактор, наряду с теоремами Геделя и Тьюринга, рассмотренными выше, дает решение 18-й проблемы С.Смейла.

Мы можем полагать, что искусственный интеллект, способный делать случайные открытия, будет представлять собой опасный агент. Кто даст гарантии, что машина, умеющая распознавать непреднамеренные находки, однажды не откроет некую технологию, которая нанесет ущерб человеку? На наш взгляд, при наличии надлежащих мер контроля такие ситуации можно полностью исключить (предотвратить). Преимущества, которые мы получаем при использовании вычислительных возможностей (ресурсов) искусственного интеллекта, весьма впечатляющие. Мы можем рассчитывать на то, что эти машинные ресурсы помогут нам в решении многих проблем, в том числе медицинских (связанных с лечением тяжелых заболеваний и поиском путей

радикального продления человеческой жизни). Уже сейчас разновидности глубоких нейросетей, то есть технологии машинного обучения, совершают настоящую революцию в различных отраслях биологии. Они участвуют в диагностике заболеваний, выявлении генных мутаций, разработке новых лекарств, создании основ персонализированного здорового питания и т.д. Эрик Тополь в книге «Искусственный интеллект в медицине» [35] отмечает: «...Даже сейчас, в наше время, соединение усилий и талантов людей с возможностями искусственного интеллекта может привести к триумфальному успеху, если человек и машина действуют синергически» [35]. «...Истинная медицина, использующая искусственный интеллект, - подчеркивает автор, - возможна – и крайне желательна. Соединение мощи машин и людей – мощи человеческого и искусственного интеллекта – поднимет уровень медицины на беспрецедентную высоту. Мы увидим, что на этом пути нас подстерегает великое множество препятствий. Путь будет нелегким и очень долгим. Но целенаправленные усилия непременно помогут достичь желаемого» [35].

Литература:

1. Вейль Г. Математика и логика // Вейль Г. Избранные труды. Математика и теоретическая физика. – М.: «Наука», 1984. - С.328-340.
2. Вейль Г. Полвека математики. – М.: «Знание», 1969. – 47 с.
3. Частиков А.П. Архитекторы компьютерного мира. – СПб.: «БХВ-Петербург», 2002. – 384 с.
4. Нагель Э., Ньюмен Дж. Теорема Геделя. – М.: «Знание», 1970. – 63 с.
5. Манин Ю.И. Теорема Геделя // Природа. – 1975. - № 12. – С.80-87.
6. Пенроуз Р. Новый ум короля. О компьютерах, мышлении и законах физики. – М.: Едиториал УРСС, 2003. – 384 с.
7. Новиков Н.Б. Решение 18-й проблемы С.Смейла. – М.: Институт психологии РАН, 2013. – 1463 с.

8. Новиков Н.Б. Каковы пределы интеллекта, как искусственного, так и человека? // Нейроинформатика. – 2013. – Том 7. - № 1. – С.42-62.
9. Новиков Н.Б. 18-я проблема С.Смейла в зеркале случайных открытий. – М.: Институт психологии РАН, 2017. – 1390 с.
10. Майданов А.С. Искусство открытия: методология и логика научного творчества. – М.: изд-во «Репро», 1993. – 175 с.
11. Азерников В. Продолжение следует // Наука и жизнь. – 1968. - № 1. – С.62-67.
12. Ходаков Ю.В. Как рождаются научные открытия. – М.: «Наука», 1964. – 96 с.
13. Александрин В.В. Падре Реанимационе // Химия и жизнь. – 2000. - № 1. – С.28-31.
14. Семенов Н.Н. Таким образом, я пришел к идее... // Химия и жизнь. – 1986. - № 4. – С.39-44.
15. Мазин В.В., Шашкова Л.С. Грибы, растения и люди. – М.: «Агропромиздат», 1986. – 208 с.
16. Сухотин А.К. Превратности научных идей. – М.: «Молодая гвардия», 1991. – 271 с.
17. Харгиттаи И. Откровенная наука. Беседы с корифеями биохимии и медицинской химии. – М.: «КомКнига», 2006. – 544 с.
18. Елдышев Ю.Н. Нобелевские премии – 2006: научная олимпиада или чемпионат США? // Экология и жизнь. – 2006. - № 12 (61). – С.38-42.
19. Дмитриев И.С. Электрон глазами химика. Очерки о современной квантовой химии. – Ленинград: «Химия», 1986. – 228 с.
20. Мигдал А.Б. Нильс Бор и квантовая физика // Успехи физических наук. – 1985. – Том 147. - № 2. – С.303-342.
21. Данин Д.С. Вероятностный мир. – М.: «Знание», 1981. – 208 с.

22. Симо К., Смейл С., Шенсине А. и др. Современные проблемы хаоса и нелинейности. – Ижевск: Институт компьютерных исследований, 2002. – 304 с.
23. Марихин В.А. Синтетические металлы // Химия и жизнь. – 2000. - № 6. – С.10-13.
24. Канер Р., Макдайрмид Э. (Мак-Диармид). Электропроводящие полимеры // В мире науки. – 1988. - № 4. – С.50-56.
25. Гинзбург В.Л. Несколько замечаний об изучении сверхпроводимости // Успехи физических наук. – 2005. – Том 175. - № 2. – С.187-190.
26. Музрукова Е.Б., Фандо Р.А. Создание хромосомной теории наследственности // Природа. – 2015. - № 7. – С.79-86.
27. Городилов Ю.Н. Организатор Шпемана: его источники и его производные (клеточно-тканевые и молекулярно-генетические аспекты) // Цитология. – 2001. – Том 43. - № 2. – С.182-203.
28. Капто А.А. Анналы урологии: от 5000 года до н.э. до 2014 года. Справочно-энциклопедическое исследование. – М.: «Полиграф-Информ», 2014. – 544 с.
29. Богданов А.А. Памяти Марианн Грюнберг-Манаго // Биохимия. – 2013. – Том 78. - № 4. – С.562-564.
30. Абрагам А. Время вспять, или Физик, физик, где ты был? – М.: «Наука», 1991. – 391 с.
31. Курбатов В.И. Логика. Систематический курс. – Ростов-на-Дону: «Феникс», 2001. – 512 с.
32. Шульпин Г.Б. Мир необычных молекул. Металлоорганические комплексы. - М.: «Наука», 1986. – 173 с.
33. Афанасьев В.А., Заиков Г.Е. В мире катализа. – М.: «Наука», 1977. – 107 с.
34. Рассел С., Норвиг П. Искусственный интеллект: современный подход. – М.: «Вильямс», 2006. – 1408 с.
35. Тополь Э. Искусственный интеллект в медицине: как умные технологии меняют подход к лечению. – М.: «Альпина Паблишер», 2022. – 400 с.