

## НАУКА, ТЕХНИКА, ОБЩЕСТВО

А. Элбакян

### Концепция неоткрытого знания Дона Р. Суонсона и искусственный интеллект

*Элбакян Александра* – руководитель проекта Sci-Hub; e-mail: alexandra@elbakyan.io

В 2024 г. исполняется 100 лет со дня рождения Дона Р. Суонсона. Работы этого ученого по теории познания в среде специалистов по философии практически неизвестны, в то время как в области искусственного интеллекта они получили широкое признание и стали основой отдельного направления – автоматической генерации научных гипотез. Задача данной статьи состоит в том, чтобы познакомить читателей с оригинальной концепцией неоткрытого знания, разработанной Д.Р. Суонсоном, и проанализировать его соотношение с другими видами знания: открытым и неявным. Развивая идеи Поппера, автор концепции утверждает, что источником нового, нетривиального знания может выступать не только физический мир, но и мир «объективного знания» – следовательно, поиск новых научных гипотез должен начинаться не в лаборатории, а в библиотеке. Такой поиск может быть автоматизирован при помощи технологий искусственного интеллекта. В статье показано, что критически важным для прогресса науки является и развитие технологий открытого доступа к научным знаниям.

**Ключевые слова:** искусственный интеллект, открытая наука, неявное знание, языковые модели, обработка естественного языка, компьютерная эпистемология, социальная эпистемология, библиометрия, научные библиотеки, медицина

### Введение

Прошедший год можно назвать годом научно-технической революции, связанной с прогрессом в технологиях искусственного интеллекта: нейронные сети научились говорить и рисовать на уровне, который сопоставим с человеческим. Как следствие, сегодня вновь происходит всплеск интереса к проблематике искусственного интеллекта в философии [Дубровский, 2023; Chalmers, 2023].

В связи с этим нельзя не заметить, что в большинстве случаев философская мысль следует за технологическим прогрессом, но не определяет его. Несмотря на то, что философские представления определяют предпосылки познания, задают необходимый фундамент деятельности как ученого, так и инноватора – они остаются в сфере личностного или неявного знания, в полной мере не осознаются субъектом или, напротив, рассматриваются как самоочевидные. Труды по философии познания, как правило, не упоминаются в списке литературы в публикациях по естественным и техническим наукам, а если и цитируются во введении, то не являются частью цепочки рассуждений автора, а только задают общий фон или некоторые рамки, в границах которых выполнено то или иное исследование.

Концепция неоткрытого знания Дона Р. Суонсона представляет собой редкое исключение из этого правила. Возможно, причиной послужило то, что автор опубликовал свою концепцию не в специализированном философском журнале, а в нескольких изданиях по информатике, библиотечному делу, медицине и биологии, в результате чего на эти работы, написанные во второй половине восьмидесятых годов, современные исследователи в области искусственного интеллекта ссылаются до сих пор, они стали основой отдельного направления в ИИ: автоматической генерации научных гипотез. С другой стороны, по той же самой причине в философии познания идеи Суонсона практически неизвестны. Данная статья призвана восполнить этот пробел и вернуть концепцию неоткрытого знания обратно в эпистемологию.

### Неоткрытое знание в научных текстах

Исходным пунктом рассуждений автора стала философия познания Карла Поппера, который сформулировал две важные идеи [Swanson, 1986]: (а) научное знание растет в результате появления новых гипотез, а не накопления фактов; (б) традиционное понимание «знания» как ментального состояния субъекта ошибочно, поскольку знание научное является объективным и автономным, это «знание без субъекта знания» [Поппер, 2002, с. 111]. Поппер утверждал, что наряду с миром идей и миром материальным существует «третий мир» объективного знания, к которому относятся научные рассуждения, теории и гипотезы, содержание журналов, книг и библиотек. Являясь продуктом человеческой мысли, объективное знание одновременно существует независимо от человека и развивается по собственным законам.

Как и в физическом мире, в мире объективного знания есть много неоткрытых мест – например, простые числа существуют до того, как о них станет известно математикам. Основываясь на этой идее Поппера, автор утверждает, что научное знание может быть “public yet undiscovered”: опубликованным, но все еще не открытым [Swanson, 1986, p. 108]. Для доказательства этого утверждения приводятся три мысленных эксперимента.

Допустим, в науке на данный момент принята гипотеза о том, что все лебеди белые. Предположим, что группа исследователей из отдаленных уголков земного шара, которым неизвестно о данной гипотезе, публикуют в местном журнале статью о группе черных лебедей. Из чего следует вывод: гипотеза о белых лебедях опровергнута, и этот результат является частью объективного,

опубликованного знания, однако об этом никому неизвестно; это знание остается неоткрытым, что связано с «трудностями в выявлении, поиске и получении доступа к соответствующей опубликованной информации» (здесь и далее перевод мой. – А.Э.) [Swanson, 1986, p. 109].

Если приведенный выше пример представляется надуманным, то следующий аргумент звучит гораздо более интересно. Допустим, группа ученых провела исследование и обнаружила, что процесс А вызывает результат В, а другая независимая группа ученых параллельно доказала утверждение «В вызывает С». Таким образом, утверждение «А вызывает С» становится частью объективного и опубликованного знания, но присутствует в нем неявно, связь между А и С является имплицитной [Swanson, 1990a]. Пример – польза рыбьего жира при болезни Рейно [Swanson, 1987]. Ряд исследований показал, что употребление рыбьего жира приводит к уменьшению вязкости крови; при этом другие исследования указывают на увеличение вязкости крови при болезни Рейно. Таким образом, анализ литературы дает возможность предположить, что применение рыбьего жира облегчит состояние пациентов, несмотря на то, что в явном виде эта гипотеза нигде не была высказана. Аналогичным образом из литературы можно получить гипотезу о дефиците магния как причине головных болей [Swanson, 1988a] и о влиянии аргинина на процессы старения [Swanson, 1990b]. При этом связь между разными идеями необязательно должна быть строго логической или причинной, к новым гипотезам могут приводить и неожиданные аналогии [Swanson, 1990a].

Третий аргумент автора звучит так: в естественных науках нередко ситуации, когда теория выдерживает множество независимых проверок, однако каждый отдельный эксперимент сам по себе не является достаточно сильным, поскольку допускает иные возможные объяснения полученных результатов; однако они исключаются, если рассматривать результаты всех проверок вместе. Пример – гипотеза о влиянии курения табака на вероятность развития раковых заболеваний, которая долгое время вызывала сомнения, поскольку статистика смертности, результаты экспериментов, взятые отдельно и сами по себе, не являлись достаточным обоснованием тезиса. Неоткрытое знание подобно кусочкам головоломки, его необходимо собрать из логически связанных, но независимых фрагментов, чтобы обнаружить «скрытую гипотезу», которая в неявном виде присутствует в массиве опубликованных научных текстов [Swanson, 1986, p. 117].

Итак, не только физический мир может выступать источником нового нетривиального знания и научных открытий, но и мир научной литературы: «новое знание можно получить как в лаборатории, так и в библиотеке» [Swanson, 1988b, p. 96]. Если же отталкиваться от идеи Поппера о первичности гипотезы в научном поиске, то библиотека имеет приоритет над лабораторией – однако эта истина не осознается в полной мере научным сообществом:

Система вознаграждения и этос науки, однако, признают только физический мир в качестве источника новых знаний. Литература, как правило, рассматривается как своего рода некролог знаний, механизм распространения, который помогает делать открытия в лаборатории, но не имеет собственной жизни. Гораздо более полезным для науки мог бы стать новый

взгляд на литературу как на огромную мозаику еще не открытых связей, и потенциальный источник бесконечного множества идей в самых разных комбинациях – мир, границы которого простираются в бесконечность [Swanson, 1990a, p. 36].

Дон Суонсон назвал свой метод поиска нового знания в библиотеке “literature-based discovery” или «открытие на основе литературы», и предвидел создание в будущем систем искусственного интеллекта, которые смогут автоматизировать научные открытия путем анализа научной литературы:

Описанные выше проблемы могут быть актуальны для работ по автоматической обработке текстов и систем, основанных на знаниях, в искусственном интеллекте. Подход к изучению литературы, изложенный в данной статье, может быть полезен при определении проблем и требований к системам будущего [Ibid., p. 35].

Термин “literature-based discovery” сегодня используется для обозначения отдельной области исследований, которая занимается разработкой систем и алгоритмов автоматического извлечения знаний из научных текстов [Bruza, Weeber, 2008].

### **Автоматизация научных открытий**

В середине 90-х гг. появились первые алгоритмы для того, чтобы извлекать из научных текстов информацию о взаимодействии различных белков, молекул, тканей, генов, заболеваний и лекарственных препаратов и т.д., но речь шла только о взаимодействиях, напрямую обозначенных в тексте, а не о решении задачи поиска неявного знания и новых гипотез.

Первые алгоритмы компьютерного поиска гипотез по методологии Суонсона [Swanson, Smalheiser, 1997; Weeber et al., 2001] были только частично автоматическими: сначала с помощью поиска по ключевым словам создавалась подборка научных публикаций, например, по теме определенного заболевания. Затем составлялся рейтинг терминов, которые наиболее часто упоминались в найденных документах. По этим терминам вновь выполнялся поиск и т.д. Результатом такой цепочки становилась подборка терминов, которые обозначали причину болезни либо лекарство от нее – они выступали «конечной точкой» исследования.

Применяя этот подход, алгоритм [Srinivasan, Libbus, 2004] выдвинул гипотезу о полезности куркумы при болезни Крона, заболеваниях глаз и позвоночника. Метод работал автоматически, но использовал метаданные MEDLINE о ключевых терминах для каждой статьи; эти метаданные создаются экспертами вручную.

Современные алгоритмы работают иначе: они основаны на представлении слов и документов в виде математических векторов, т.е. последовательности чисел [Раскатова, Чельшев, 2022], при этом похожие слова и документы конвертируются в похожие числовые векторы. В самом простом случае такой вектор получается путем подсчета частоты употребления каждого слова в документе либо в контексте другого слова. Поскольку близкие по семантике слова употребляются

в похожих контекстах, то они имеют похожие частотные векторы. На этой идее основан алгоритм [Spangler et al., 2014]. Проанализировав порядка 70 000 аннотаций к научным статьям, суперкомпьютер IBM смог обнаружить несколько ранее неизвестных ферментов, которые взаимодействуют с белком p53 – этот белок играет в организме человека важную роль при предотвращении развития раковых опухолей [Чумаков, 2007]. Поскольку похожие по своим функциям ферменты употребляются в научных текстах в близком контексте, то, опираясь на список уже известных науке ферментов, которые взаимодействуют с p53, алгоритм смог отыскать новые ферменты с аналогичными функциями. Если говорить словами авторов, то алгоритм «автоматически обнаружил скрытое сходство между объектами на основе корпуса научных статей» [Spangler et al., 2014, p. 1878]. Несмотря на достаточно большую дистанцию, которая отделяет это исследование от более ранних попыток, концепция Д. Суонсона упоминается и здесь в качестве основополагающей для такого направления, как компьютерная генерация гипотез.

В современных алгоритмах для получения числового вектора каждого слова используются нейронные сети, но общий принцип работы остается прежним. В исследовании [Tshitoyan et al., 2019] нейронная сеть была обучена на корпусе текстов, который состоял из трех миллионов аннотаций к научным статьям из области материаловедения. В результате были открыты новые материалы с термоэлектрическими свойствами: названия таких материалов имели векторы, близкие к числовому вектору слова «термоэлектрик», из чего авторы сделали вывод, что алгоритм имеет потенциал «открыть скрытое знание, которое недоступно ученым напрямую» [Ibid., p. 97].

Если в начале для вычисления векторов использовались простые нейронные сети, то сегодня их сменили многослойные или «глубокие» нейросети, к которым относятся широко известные языковые модели BERT и GPT. Есть много специализированных вариантов BERT для научных текстов, таких как BioBERT и SciBERT; последняя была обучена на полных текстах примерно одного миллиона научных статей, 80% которых составляли статьи по биологии и медицине, а еще 20% приходилось на статьи по биоинформатике [Beltagy et al., 2019]. Вариаций на основе других моделей сравнительно меньше, они только начинают появляться: в качестве примеров можно назвать AstroLLaMA [Nguyen et al., web], OceanGPT [Bi et al., web], Radiology-GPT [Liu et al., web]. Специальные модели работают точнее: например, AstroLLaMA правильно определяла научные статьи по аналогичным темам там, где обученный на текстах общего характера ChatGPT ошибался.

Несмотря на большой поток публикаций, посвященных обучению нейросетей на научных текстах, тематика извлечения неявного знания в них не затрагивается, но именно в этом направлении очевиден огромный потенциал языковых моделей: поскольку они могут самостоятельно генерировать тексты, то экспликация неявного научного знания могла бы осуществляться напрямую, не ограничиваясь простым сравнением числовых векторов отдельных научных терминов, за счет «генерирования текстов, касающихся определенной предметной области и соответствующих заданным эпистемическим параметрам» [Алексеева, 2023, с. 97].

## Неоткрытое знание в гуманитарных науках

Концепция неоткрытого знания была разработана для естественно-научных текстов; что касается применения метода Д. Суонсона в гуманитарных науках, то согласно К.А. Кори «извлекать скрытое знание из гуманитарных баз данных особенно проблематично» [Coqu, 1997, p. 60]. Во-первых, необходимо работать с текстами, написанными обыденным, «повседневным» языком, который недостаточно точен, в отличие от текстов медицинской тематики, где применяются специальные термины; во-вторых, метод гуманитарных наук подразумевает поиск аналогий, а не причинно-следственных связей.

Тем не менее поиск неоткрытого знания может работать и в гуманитарных науках, но это требует изменения подхода: вместо специальных терминов необходимо использовать фамилии авторов при формулировке поискового запроса. Кроме того, значимыми являются не наиболее частотные термины, как при анализе медицинских гипотез, а любая фамилия, даже если она упоминается всего один раз.

К. Кори продемонстрировал работу метода, выполнив поиск по ключевым словам «Джеймс и эпистемология или прагматизм». Запрос вернул порядка 80 статей, из которых две статьи, сопоставленные вместе, дали новый нетривиальный результат: первая работа была посвящена параллелям между эпистемологическими идеями Джеймса и древнегреческого философа Карнеада, а во второй говорилось о влиянии философии Джеймса на поэзию Р. Фроста – это позволило выдвинуть гипотезу, согласно которой идеи античного философа могут быть ключом к пониманию и интерпретации его стихотворений.

Исследование К. Кори пока остается единственным по данной теме; нет и специальных языковых моделей, а производительность нейросетей общего характера в гуманитарных науках низкая, как показало исследование пяти популярных чат-ботов: в эксперименте они должны были ответить на дискуссионный вопрос по истории миграции славянских племен; эта тематика находится на стыке нескольких дисциплин: археологии, лингвистики, историографии и генетики. Ответы оценивались как на предмет их корректности и соответствия последним научным данным, так и на предмет оригинальности. Наиболее качественный ответ дал ChatGPT, но и его уровень был ниже того, который можно было бы ожидать от студента старшего курса. При этом оценка оригинального научного вклада чат-ботов не превышала 11%, в то время как для людей этот показатель достигал 50% и более, из чего авторы сделали вывод о том, что чат-боты, «вероятно, никогда не смогут генерировать новое знание подобно человеку» [Lozić, Štular, 2023, p. 19].

## Неоткрытое и неявное знание

В англоязычной литературе для обозначения «неоткрытого» знания используются термины “hidden” скрытое, “latent” латентное, “implicit” имплицитное; последние два термина также переводятся на русский язык как «неявное».

В эпистемологии существует обширная традиция исследования неявного знания, родоначальником которой является М. Полани: он определял неявное знание как такое, которое не высказано или не может быть высказано словами

[Микешина, Опенков, 1998]. Наиболее простым примером неявного знания Полани считал восприятие, способность узнать человека в лицо. При этом идеал современной науки составляет именно эксплицитное или явное знание.

Важное отличие здесь состоит в том, что неявное знание часто отождествляется с интуитивным и противопоставляется рациональному, в то время как неоткрытое знание представляет собой вполне рациональный логический вывод; к такому виду знания можно отнести все открытия в математике, например аксиомы в геометрии в неявной форме содержат в себе все логически выводимые из них теоремы.

Можно провести интересные параллели между концепцией неоткрытого знания и диалектикой явного и скрытого в арабской философии. Арабские мыслители рассматривали текст Корана как содержащий в себе скрытые смыслы: они могут быть выявлены путем толкования. Также «непосредственное» знание, данное избранным людям свыше, может быть истолковано, то есть сообщено другим людям в дискурсивной форме. Здесь два вида знания, непосредственное и дискурсивное, предстают как целое и его часть: «как в яйце заключается целиком вся птица, так целокупное заключает в себе частичное в качестве скрытого, а потому нуждается в чем-то, что это скрытое выявит» [Смирнов, 2013, с. 144–145].

Более того, долгое время арабский Восток являлся центром алхимических изысканий; известно, что алхимики применяли различные методы шифрования своих манускриптов. Так называемый «метод дисперсии» заключался в том, что учение не излагалось последовательно в одном тексте, а делилось на кусочки, которые распределялись по разным трактатам: читатель должен был самостоятельно их собрать. Таким образом, процесс извлечения скрытого знания из библиотеки алхимических манускриптов практически идентичен процессу поиска «скрытых гипотез» в базе данных научной литературы.

Интересно, что сама идея поиска скрытых гипотез появилась у Д. Суонсона во время работы над проектом для военной разведки:

Считается, что новое знание, или итоговое заключение разведки, возникает из большого количества тщательно собранных деталей [фрагментов данных], каждая из которых по отдельности не важна и связь между которыми изначально не была очевидна [Swanson, 1988b, p. 95].

Автор заметил, что подход разведчика отличается от подхода ученого: последний основным источником знания видит физический мир, а не уже собранную базу данных научного знания, которая доступна в библиотеке, поэтому науке следует перенимать у разведывательного сообщества культуру работы с информацией. Несмотря на то, что автор верно определяет проблему, с данным выводом сложно согласиться, поскольку в разведсообществе одной из главных ценностей является секретность, что полностью противоречит ценностям научного сообщества, где принята открытая коммуникация; этот вопрос будет подробнее рассмотрен далее.

К сказанному выше можно только добавить, что, поскольку отдельные фрагменты неоткрытого знания находятся в работах большого числа разных

исследователей, то такое знание является по своей природе коллективным; один из основателей социальной эпистемологии Стив Фуллер называл Дона Суонсона «настоящим пионером» этой области [Fuller, 2023, p. 64].

### Неоткрытое и открытое знание

Во время холодной войны на страницах научных журналов развернулась дискуссия, связанная с проблемой ограниченного доступа к информации. Большинство исследователей выступало против режима секретности в науке, аргументируя свою позицию тем, что открытая коммуникация является необходимым условием научного прогресса. Так, создатель водородной бомбы Э. Теллер писал, что секретность несовместима с наукой, порождает невежество и препятствует прогрессу [Teller, 1979]. В конце XX в. дебаты об открытом доступе к знаниям возобновляются в ином контексте: коммерциализация научных журналов привела к появлению финансовых барьеров, ограничивающих доступ к научному знанию не только для широкой публики, но и для многих ученых [Guédon, 2003].

Таким образом, сегодня можно говорить о существовании понятия «открытого знания» как такого знания, доступ к которому не ограничен, однако по сравнению с неявным открытое знание в эпистемологии остается практически неразработанной темой; «Социология невежества» [Штейнзалц, Функенштейн, 1997] является одной из немногих работ в данном ключе. Интересно, что деление наук на отдельные дисциплины авторы работы рассматривают как форму закрытости: «такого рода знание можно назвать закрытым по вертикали, то есть закрытым перед людьми других профессий или относительно иных специализаций» [Там же, с. 90]. Здесь можно провести параллель с концепцией Д. Суонсона, который полагал, что основное количество неоткрытого знания является результатом неоднородности пространства научных публикаций: оно состоит из отдельных «библиометрически изолированных» островков или кластеров [Swanson, 1990a, p. 29]. Авторы внутри каждого кластера активно цитируют друг друга, в то время как взаимные цитирования между кластерами практически отсутствуют. Так, научные публикации по двум темам – болезни Рейно и рыбьего жира – формируют отдельные изолированные группы. Данные о влиянии рыбьего жира на состав крови человека содержатся в одном «библиометрическом кластере», а информация об особенностях состава крови при болезни Рейно – в другом. Прийти к выводу, что рыбий жир полезен при болезни Рейно, трудно из-за того, что отдельные фрагменты гипотезы находятся в разных пространствах, «разбросаны и утоплены в море литературы» [Ibid., p. 32], поэтому знание о пользе рыбьего жира при болезни Рейно остается неоткрытым, существует в неявной форме, и его обнаружение является нетривиальной задачей. Таким образом, неоткрытое знание является результатом специализации, дробления единого научного знания на отдельные дисциплины.

Можно отметить, что проблема становится гораздо более серьезной при наличии финансовых барьеров: в этом случае библиотеки университетов, как правило, выписывают научные журналы в ограниченном количестве только

по своей специальности, поскольку они вынуждены экономно расходовать средства.

Развивая эту мысль, Д. Суонсон приходит к выводу, что новое знание получается в результате соединения фрагментов, которые до этого находились буквально на отдельных «островах». Похожая идея встречается у Л. Беркнера: он аргументировал, что секретность вредна для прогресса, поскольку новые идеи в науке появляются в результате неожиданного соединения разнообразных фактов и идей, которые раньше казались несвязанными. Например, Эйнштейн объединил идеи Планка с концепциями механики и астрономии, но не смог бы сделать этого сегодня, поскольку работа Планка по текущим правилам обязательно оказалась бы засекреченной [Berkner, 1956]. Таким образом, новое открытие в науке – это соединение несвязанного, причем новизна открытия тем больше, чем более отдаленные фрагменты научного знания оно соединяет.

### **Технологии открытого знания**

Итак, неоткрытое знание возникает в тех случаях, когда часть информации недоступна или не может быть найдена, но основной акцент в своих работах Д. Суонсон делает на проблеме поиска. Например, опровержение гипотезы о том, что все лебеди белые, остается неоткрытым знанием потому, что небольшой местный журнал, в котором была опубликована статья о группе черных лебедей, трудно найти, поэтому разработка более эффективных технологий поиска необходима для раскрытия неоткрытого знания.

Вместе с тем можно заметить, что различие между поиском и доступом достаточно условно: о местном журнале можно сказать, что он труднодоступен. Возможно, проблему неоткрытого знания правильно будет рассматривать не как проблему эффективного поиска информации, а именно как проблему доступа к информации – тогда большее значение для прогресса науки имеет разработка не поисковых технологий, а инновационных технологий открытого доступа к научному знанию. Примером такой технологии является сервис Sci-Hub, который предоставляет открытый доступ к базе данных научной литературы объемом около 88 миллионов документов; это достаточно большая база данных, а «эмерджентное свойство очень больших баз данных литературы», согласно Д. Суонсону, состоит в том, что объем неоткрытого знания в них многократно превышает количество тех научных фактов, которые высказаны в явном виде, поскольку с добавлением каждого нового документа количество возможных связей между отдельными фактами растет нелинейно [Swanson, 1990a, p. 35]. Таким образом, обучение искусственного интеллекта на такой огромной базе данных научной литературы, как у проекта Sci-Hub, может привести к совершенно неожиданным результатам.

### **Заключение**

Дон Р. Суонсон родился 10 октября 1924 г. и получил ученую степень в области теоретической физики, но основной вклад сделал в области теории познания. В 2024 г. исполняется 100 лет со дня рождения ученого: этот

юбилей можно отметить признанием его оригинальной эпистемологической концепции, которая позволяет по-новому посмотреть на многие вопросы о научном познании, природе интуиции и научного открытия, а также оценить потенциальные возможности и будущую роль искусственного интеллекта в развитии человечества. Роль эта может оказаться исключительно положительной: развитие технологий искусственного интеллекта позволит автоматизировать процесс научного поиска в таких важных областях, как биология и медицина, что позволит разрабатывать эффективные методы лечения различных заболеваний.

Автоматизация поиска научных гипотез особенно интересна потому, что способность выдвигать новые интересные идеи и гипотезы относится к творческим компонентам научной деятельности; принято считать, что в этом процессе ведущая роль принадлежит интуиции и воображению – исключительно человеческим способностям, автоматизация которых до недавнего времени считалась в принципе невозможной.

### Список литературы

Алексеева, 2023 – *Алексеева Е.А.* Проекты компьютерной эпистемологии // *Философия науки и техники*. 2023. Т. 28. № 2. С. 88–101.

Дубровский, 2023 – *Дубровский Д.И.* Коммуникативный подход к анализу критериев оценки социальной значимости развития искусственного интеллекта // *Философия науки и техники*. 2023. Т. 28. № 2. С. 79–87.

Микешина, Опенков, 1998 – *Микешина Л.А., Опенков М.Ю.* Знание неявное // *Культурология. XX век. Энциклопедия: в 2 т. Т. 1 / Под ред. С.Я. Левит. М.: Университетская книга, 1998. С. 224–228.*

Поппер, 2002 – *Поппер К.Р.* Объективное знание. Эволюционный подход / Пер. с англ. Д.Г. Лахути. М.: УРСС, 2002. 384 с.

Раскатова, Чельшев, 2022 – *Раскатова М.В., Чельшев Э.А.* Векторизация текстов в задачах обработки естественного языка: история и развитие // *Современное программирование: материалы IV Международной научно-практической конференции (г. Нижневартовск, 8 декабря 2021 года) / Отв. ред. Т.Б. Казиахмедов. Нижневартовск: Нижневартовский государственный университет, 2022. С. 284–288.*

Смирнов, 2013 – *Смирнов А.В.* История арабо-мусульманской философии: Учебник. М.: Академический Проект, 2013. 255 с.

Чумаков, 2007 – *Чумаков П.М.* Белок р53 и его универсальные функции в многоклеточном организме // *Успехи биологической химии*. 2007. Т. 47. № 1. С. 3–52.

Штейнзальц, Функенштейн, 1997 – *Штейнзальц А., Функенштейн А.* Социология нееврейства. М.: Институт изучения иудаизма в СНГ, 1997. 166 с.

Beltagy, Lo, Cohan, 2019 – *Beltagy I., Lo K., Cohan A.* SciBERT: A Pretrained Language Model for Scientific Text // *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9<sup>th</sup> International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 2019. P. 3615–3620.

Berkner, 1956 – *Berkner L.V.* Secrecy and scientific progress // *Science*. 1956. Vol. 123. № 3201. P. 783–786.

Bi et al., web – *Bi Z. et al.* OceanGPT: A Large Language Model for Ocean Science Tasks // *arXiv*. 2023. URL: <https://arxiv.org/abs/2310.02031> (accessed on: 19.07.2024).

Bruza, Weeber, 2008 – *Bruza P., Weeber M.* (eds.) *Literature-based discovery*. Heidelberg: Springer, 2008. 197 p.

- Chalmers, 2023 – *Chalmers D.J.* Could a large language model be conscious? // arXiv. 2023. URL: <https://arxiv.org/abs/2303.07103> (accessed on: 19.07.2024).
- Cory, 1997 – *Cory K.A.* Discovering hidden analogies in an online humanities database // *Computers and the Humanities*. 1997. Vol. 31. № 1. P. 1–12.
- Fuller, 2023 – *Fuller S.A.* World Information Strategy for the Future // *Computer Sciences & Mathematics Forum*. 2023. Vol. 8. № 1. P. 62–66.
- Guédon, 2003 – *Guédon J.-C.* Open Access Archives: from scientific plutocracy to the republic of science // *IFLA Journal*. 2003. Vol. 29. № 2. P. 129–140.
- Liu et al., web – *Liu Z. et al.* Radiology-GPT: A Large Language Model for Radiology // arXiv. 2023. URL: <https://arxiv.org/abs/2306.08666> (accessed on: 19.07.2024).
- Lozić, Štular, 2023 – *Lozić E., Štular B.* Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots' Proficiency and Originality in Scientific Writing for Humanities // *Future Internet*. 2023. Vol. 15. № 10. P. 336.
- Nguyen et al., web – *Nguyen T.D. et al.* AstroLLaMA: Towards Specialized Foundation Models in Astronomy // arXiv. 2023. URL: <https://arxiv.org/abs/2309.06126> (accessed on: 19.07.2024).
- Spangler et al., 2014 – *Spangler S. et al.* Automated hypothesis generation based on mining scientific literature // *Proceedings of the 20<sup>th</sup> ACM SIGKDD international conference on Knowledge discovery and data mining*. 2014. P. 1877–1886.
- Srinivasan, Libbus, 2004 – *Srinivasan P., Libbus B.* Mining MEDLINE for implicit links between dietary substances and diseases // *Bioinformatics*. 2004. Vol. 20. No. suppl\_1. P. i290–i296.
- Swanson, 1986 – *Swanson D.R.* Undiscovered public knowledge // *The Library Quarterly*. 1986. Vol. 56. № 2. P. 103–118.
- Swanson, 1987 – *Swanson D.R.* Two medical literatures that are logically but not bibliographically connected // *Journal of the American Society for Information Science*. 1987. Vol. 38. № 4. P. 228–233.
- Swanson, 1988a – *Swanson D.R.* Migraine and magnesium: eleven neglected connections // *Perspectives in biology and medicine*. 1988. Vol. 31. № 4. P. 526–557.
- Swanson, 1988b – *Swanson D.R. et al.* Historical note: Information retrieval and the future of an illusion // *Journal of the American Society for Information Science*. 1988. Vol. 39. № 2. P. 92–98.
- Swanson, 1990a – *Swanson D.R.* Medical literature as a potential source of new knowledge // *Bulletin of the Medical Library Association*. 1990. Vol. 78. № 1. P. 29–37.
- Swanson, 1990b – *Swanson D.R.* Somatostatin C and arginine: implicit connections between mutually isolated literatures // *Perspectives in biology and medicine*. 1990. Vol. 33. № 2. P. 157–186.
- Swanson, Smalheiser, 1997 – *Swanson D.R., Smalheiser N.R.* An interactive system for finding complementary literatures: a stimulus to scientific discovery // *Artificial intelligence*. 1997. Vol. 91. № 2. P. 183–203.
- Teller, 1979 – *Teller E.* *Energy from Heaven and Earth*. San Francisco: W.H. Freeman, 1979. 322 p.
- Tshitoyan et al., 2019 – *Tshitoyan V. et al.* Unsupervised word embeddings capture latent knowledge from materials science literature // *Nature*. 2019. Vol. 571. № 7763. P. 95–98.
- Weeber et al., 2001 – *Weeber M. et al.* Using concepts in literature-based discovery: Simulating Swanson's Raynaud-fish oil and migraine-magnesium discoveries // *Journal of the American society for information science and technology*. 2001. Vol. 52. No. 7. P. 548–557.

## Don R. Swanson's concept of undiscovered knowledge and artificial intelligence

*Alexandra Elbakyan*

Sci-Hub; e-mail: alexandra@elbakyan.io

The 2024 will be an anniversary year for Don R. Swanson, who was born 100 years ago. The works of Swanson are almost unknown in philosophy, yet his name is well recognized in the field of artificial intelligence, where it became foundational for a new sub-field of automatic generation of scientific hypotheses. The aim of this article is to introduce readers to the original concept of undiscovered knowledge proposed by D.R. Swanson. The second aim will be to investigate its relationship with other types: tacit, or implicit and open knowledge. The author develops ideas of Popper by proposing that not only physical world, but also the world of objective knowledge can be a source of new, non-trivial knowledge. Therefore, the search for novel scientific hypotheses has to start in the library rather than in the lab. This search can be made automatic given advances in the technologies of artificial intelligence. The paper overviews historical development and current capabilities of AI to discover hidden knowledge in scientific literature. However, as will be shown in this article, advances in technologies of open access to knowledge is critical for the progress of science as well. This conclusion follows from the fact that undiscovered knowledge is in many respects the problem of access.

**Keywords:** artificial intelligence, open science, implicit knowledge, large language models, natural language processing, social epistemology, bibliometrics, automatic hypothesis generation, literature-based discovery, medicine

### References

- Alekseeva, E.A. "Proekty komp'yuternoj epistemologii" [Computational epistemology projects], *Filosofiya nauki i tekhniki* [Philosophy of Science and Technology], 2023, vol. 28, no. 2, pp. 88–101. (In Russian)
- Beltagy, I., Lo, K., Cohan, A. "SciBERT: A Pretrained Language Model for Scientific Text", *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019. P. 3615–3620.
- Berkner, L.V. "Secrecy and scientific progress", *Science*, 1956, vol. 123, no. 3201, pp. 783–786.
- Bi, Z., et al. "OceanGPT: A Large Language Model for Ocean Science Tasks", *arXiv*, 2023. URL: <https://arxiv.org/abs/2310.02031> (accessed on: 19.07.2024).
- Bruza, P., Weeber, M. (eds.) *Literature-based discovery*. Heidelberg: Springer, 2008. 197 pp.
- Chalmers, D.J. "Could a large language model be conscious?", *arXiv*, 2023. URL: <https://arxiv.org/abs/2303.07103> (accessed on: 19.07.2024).
- Chumakov, P.M. "Belok r53 i ego universal'nye funkcii v mnogokletochnom organizme" [The p53 protein and its universal functions in the multicellular organism], *Uspekhi biologicheskoy himii* [Advances in biological chemistry], 2007, vol. 47, no. 1, pp. 3–52. (In Russian)
- Cory, K.A. "Discovering hidden analogies in an online humanities database", *Computers and the Humanities*, 1997, vol. 31, no. 1, pp. 1–12.
- Dubrovsky, D.I. "Kommunikativnyj podhod k analizu kriteriev ocenki social'noj znachimosti razvitiya iskusstvennogo intellekta" [Communicative approach to the analysis of the criteria for assessing the social significance of artificial intelligence development], *Filosofiya nauki i tekhniki* [Philosophy of Science and Technology], 2023, vol. 28, no. 2, pp. 79–87. (In Russian)

Fuller, S.A. “World Information Strategy for the Future”, *Computer Sciences & Mathematics Forum*, 2023, vol. 8, no. 1, pp. 62–66.

Guédon, J.-C. “Open Access Archives: from scientific plutocracy to the republic of science”, *IFLA Journal*, 2003, vol. 29, no. 2, pp. 129–140.

Liu, Z., et al. “Radiology-GPT: A Large Language Model for Radiology”, *arXiv*, 2023. URL: <https://arxiv.org/abs/2306.08666> (accessed on: 19.07.2024).

Lozić, E., Štular, B. “Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots’ Proficiency and Originality in Scientific Writing for Humanities”, *Future Internet*, 2023, vol. 15, no. 10, pp. 336.

Mikeshina, L.A., Openkov, M.U. “Znanie neyavnoe” [Tacit knowledge], in: *Kul'turologiya. XX vek. Enciklopediya: v 2 t. T. 1* [Culturology. XX century. Encyclopedia: in 2 vols. Vol. 1], ed. by S.Ya. Levit. Moscow: Universitetskaya kniga Publ., 1998, pp. 224–228. (In Russian)

Nguyen, T.D. et al. “AstroLLaMA: Towards Specialized Foundation Models in Astronomy”, *arXiv*, 2023. URL: <https://arxiv.org/abs/2309.06126> (accessed on: 19.07.2024).

Popper, K.R. *Ob'ektivnoe znanie. Evolyucionnyj podhod* [Objective knowledge. Evolutionary approach], trans. from English by D.G. Lahuti. Moscow: URSS Publ., 2002. 384 pp. (In Russian)

Raskatova, M.V., Chelyshev, E.A. “Vektorizaciya tekstov v zadachah obrabotki estestvennogo yazyka: istoriya i razvitiye” [Text vectorisation in natural language processing tasks: history and development], in: *Sovremennoe programmirovaniye: materialy IV Mezhdunarodnoi nauchno-prakticheskoi konferentsii (g. Nizhnevartovsk, 08 dekabrya 2021 goda)* [Modern programming: materials of the IV International Scientific and Practical Conference (Nizhnevartovsk, December 08, 2021)], ed. by T.B. Kaziakhmedov. Nizhnevartovsk: Nizhnevartovsk State Univ. Publ., 2022, pp. 284–288. (In Russian)

Smirnov, A.V. *Istoriya arabo-musul'manskoj filosofii: Uchebnik* [The history of arabo-islamic philosophy: a coursebook]. Moscow: Akademicheskij Proekt Publ., 2013. 255 pp. (In Russian)

Spangler, S. et al. “Automated hypothesis generation based on mining scientific literature”, *Proceedings of the 20<sup>th</sup> ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 1877–1886.

Srinivasan, P., Libbus, B. “Mining MEDLINE for implicit links between dietary substances and diseases”, *Bioinformatics*, 2004, vol. 20, no. suppl\_1, pp. i290–i296.

Steinsaltz, A., Funkenstein, A. *Sociologiya nevezhestva* [The Sociology of Ignorance]. Moscow: Institut izucheniya iudaizma v SNG Publ., 1997. 166 pp. (In Russian)

Swanson, D.R. “Medical literature as a potential source of new knowledge”, *Bulletin of the Medical Library Association*, 1990a, vol. 78, no. 1, pp. 29–37.

Swanson, D.R. “Migraine and magnesium: eleven neglected connections”, *Perspectives in biology and medicine*, 1988a, vol. 31, no. 4, pp. 526–557.

Swanson, D.R. “Somatomedin C and arginine: implicit connections between mutually isolated literatures”, *Perspectives in biology and medicine*, 1990b, vol. 33, no. 2, pp. 157–186.

Swanson, D.R. “Two medical literatures that are logically but not bibliographically connected”, *Journal of the American Society for Information Science*, 1987, vol. 38, no. 4, pp. 228–233.

Swanson, D.R. “Undiscovered public knowledge”, *The Library Quarterly*, 1986, vol. 56, no. 2, pp. 103–118.

Swanson, D.R. et al. “Historical note: Information retrieval and the future of an illusion”, *Journal of the American Society for Information Science*, 1988b, vol. 39, no. 2, pp. 92–98.

Swanson, D.R., Smalheiser, N.R. “An interactive system for finding complementary literatures: a stimulus to scientific discovery”, *Artificial intelligence*, 1997, vol. 91, no. 2, pp. 183–203.

Teller, E. *Energy from Heaven and Earth*. San Francisco: W.H. Freeman, 1979. 322 pp.

Tshitoyan, V. et al. “Unsupervised word embeddings capture latent knowledge from materials science literature”, *Nature*, 2019, vol. 571, no. 7763, pp. 95–98.

Weeber, M. et al. “Using concepts in literature-based discovery: Simulating Swanson’s Raynaud-fish oil and migraine-magnesium discoveries”, *Journal of the American Society for Information Science and Technology*, 2001, vol. 52, no. 7, pp. 548–557.