УДК 004.4'414 doi: 10.15827/0236-235X.149.017-026 2025. Т. 38. № 1. С. 17–26

Использование семантического анализа для автоматизированного выявления хайповых технологий

И.В. Логинова [1]✉, А.С. Пиекалнитс [1], А.В. Соколов [1]

[1] Институт статистических исследований и экономики знаний НИУ «Высшая школа экономики», г. Москва, 101000, Россия

Ссылка для цитирования
Логинова И.В., Пиекалнитс А.С., Соколов А.В. Использование семантического анализа для автоматизированного выявления хайповых технологий // Программные продукты и системы. 2025. Т. 38. № 1. С. 17–26. doi: 10.15827/0236-235X.149.017-026
Информация о статье
Группа специальностей ВАК: 1.2.1
Поступила в редакцию: 12.07.2024 После доработки: 23.07.2024 Принята к публикации: 30.07.2024

Аннотация. Предметом проведенного исследования являются завышенные общественные ожидания от новых технологий (хайпов). В статье представлены результаты разработки и апробации методики автоматизированного выявления хайпов среди технологических тематик на основе их текстового следа на примере сферы цифровых технологий. Количество новых технологических разработок в мире постоянно растет, однако реальный потенциал их практического применения может сильно различаться. Важно понимание надежных факторов, позволяющих отличать тренды от хайпов. К числу возможных признаков хайпов отраслевые и технологические эксперты предлагают относить, как правило, отсутствие устойчивой бизнес-модели, несформированный или заведомо ограниченный потребительский рынок, большое количество более эффективных альтернатив. Выявление хайпов в технологической повестке было и остается сложной аналитической задачей. Это обусловлено несогласованностью терминологического аппарата, экспертной природой задачи, недостаточно развитыми методологическими подходами, отсутствием конкретных технических инструментов. Описываемая в статье методика включает извлечение методами обработки естественного языка и компьютерной лингвистики из нескольких десятков миллионов текстовых документов различных типов (научные публикации, патенты, рыночная аналитика) терминов, означающих технологии, расчет объективной меры хайповости каждой технологии, а также построение визуальной карты – наглядного представления ландшафта технологий, позволяющего отделить устойчивые тренды от потенциальных хайпов (тематик с высоким значением показателя хайповости). Хайп-карты в комплексе с другими аналитическими результатами могут использовать лица, принимающие решения, в задачах выявления приоритетных направлений развития, анализа текущих и прогнозирования будущих трендов, риск-менеджмента.

Ключевые слова: семантический анализ, большие данные, анализ трендов, технологическая аналитика, хайпы, оценка рисков

Введение. Потребность в выработке прикладных решений, направленных на автоматизацию задач технологической и стратегической аналитики, обусловлена рядом вызовов. Риски управленческих ошибок при стратегически значимом выборе велики, поэтому аналитика данных как процесс информационного обеспечения принятия стратегических решений должна опираться на полные и достоверные источники сведений, носить объективный и доказательный характер, а обеспечение информацией лица, принимающего решения, должно быть своевременным и релевантным (сфокусированным). Значительное влияние на деятельность технологических компаний, в частности, на формирование их научно-технологических и продуктово-рыночных стратегий, инвестиционные решения, управление ресурсами, оказывают хайпы, что повышает потребность в разработке надежных подходов для их выявления.

Хайп в контексте данной статьи означает повышенный ажиотаж к определенной теме, продукту, явлению, событию и др. [1]. Нередко в литературе под хайпом понимается избыточный, чрезмерно высокий интерес к чему-либо, за которым так или иначе следует снижение шумихи, а часто – разочарование [2, 3]. Данный феномен особенно ярко выражен в сфере технологий, где инновации проходят через несколько фаз хайпа, характеризующихся первоначальным повышенным вниманием, последующим снижением ажиотажа и окончательной стабилизацией – восстановлением или падением интереса.

Концепция хайп-цикла, или хайп-кривой, разработанная компанией Gartner, иллюстрирует такую динамику, показывая, как технологии проходят через определенные стадии хайпа: триггер инновации, пик завышенных ожиданий, дно разочарований, склон просветления и плато продуктивности (https://www.gartner.com/en/research/methodologies/gartner-hype-cycle). В литературе выделяется ряд ограничений хайп-цикла как инструмента оценки технологической зрелости, обусловленных методологическими изъянами [4], однако критики модели Gartner замечают, что сама концепция хайпа остается перспективной для применения в задачах технологического прогнозирования [4, 5]. Кроме того, хайп-цикл как продукт является источником данных для того, чтобы сделать технологическое прогнозирование более контролируемым, повторяемым и проверяемым, что подтверждается экспериментами по обучению модели машинного обучения (модели ML – *Machine Learning*) для решения задачи предсказания потенциально разрушительных технологий [6].

В настоящей статье хайп рассматривается как первая стадия формирования устойчивого во времени тренда, являющего собой накопленный концентрированный общественный запрос на решение какой-то конкретной задачи (https://lab-w.com/method_meetup/trendwatching). При этом хайп проявляет себя посредством различных сигналов, часть из которых характеризуются изначально высокой популярностью в результате высокого резонанса обсуждений в СМИ.

Возможности автоматизации аналитической задачи по выявлению хайпов часто оцениваются критически из-за необходимости учитывать контекстуальные и психологические факторы данного явления. Частичная автоматизация задачи выявления хайпов подразумевает определение хайпа в терминах фактической высокой упоминаемости явления в СМИ/новостных источниках на фоне умеренной или низкой представленности этого же явления в научно-технической документации. Таким образом, кроме классических методов автоматизации (в первую очередь машинное обучение, предиктивные модели, временные ряды и др.) применительно к выявлению хайпов, необходимо подключать аналитику больших данных из специфичных для этой цели источников текстов (прежде всего социальных сетей, но также новостных сайтов, блогов, поисковых систем и т.п.), сентимент-анализ как инструмент выявления преобладающего общественного мнения. Такой комплексный инструментарий, тем не менее, будет направлен на реализацию аналитического сценария, в рамках которого исследователь работает с некоторым перечнем гипотез относительно хайпов и по результатам объективного анализа больших текстовых данных принимает решение по каждой технологии, к которой она тяготеет, – к хайпу или к обоснованной разработке. Таким образом, автоматический инструментарий, основанный на анализе текстового следа (текстовой статистики) направлений, нацелен на получение объективных оценок, которые могут быть использованы в качестве информационно-аналитической поддержки решения данной задачи. Однако эффективный подход требует сочетания автоматизированных систем с экспертным анализом для получения более точных результатов.

Обзор существующих подходов

Особое влияние на проблематику выявления и анализа хайпов оказывает стремительное развитие технологий *искусственного интеллекта* (ИИ). Современные генеративные языковые модели могут создавать правдоподобные синтетические тексты, которые способны влиять на общественное мнение, несмотря на свое искусственное происхождение [7]. В некоторых работах отмечается, что чат-боты, использующие передовые ИИ-технологии, могут применяться для распространения дезинформации (например, генерировать фальшивые новостные статьи, выдавая их за написанные авторитетными источниками) с минимальными затратами. Это способствует усилению эффекта хайпа вокруг технологий, оказывая негативное влияние на результаты аналитики [8].

К рискам, связанным с ИИ и использованием хайпов в задачах технологического прогнозирования, относятся появление и распространение фейков и дипфейков, применяющих такие технологии, как, например, генеративно-состязательные сети для создания реалистичного, но сфабрикованного текстового, аудио- и видеоконтента. Исследования показывают, что фейки в обозримом будущем будут оказывать воздействие, подрывающее доверие к информационным источникам, и стимулировать распространение ложных нарративов [9]. При этом современные технологии их распознавания все еще находятся на недостаточном уровне развития и требуют инвестиций в большее число фундаментальных исследований в данной области, хотя некоторые из предлагаемых методологий и достигают относительно высокой точности [10–12]. Таким образом, в условиях бурного развития технологий генеративного ИИ использование хайпов как источников данных для технологического прогнозирования и стратегической аналитики может искажать результаты и приводить к ошибочным выводам из-за наличия в данных отчасти или полностью сфальсифицированных элементов.

Поскольку хайпы зачастую связаны с неоправданно завышенными ожиданиями и подвергаются искажению за счет использования технологий ИИ, актуальна задача выявления потенциальных хайпов при проведении технологической аналитики и исключения таких хайповых тем при принятии стратегических решений. Для ее решения могут применяться как количественные, так и качественные инструменты, наиболее известным из которых является ранее упомянутый хайп-цикл Gartner, позволяющий разграничивать технологии, находящиеся на пике завышенных ожиданий и менее подверженные влиянию хайпа. Однако в силу того, что экспертная методология Gartner имеет свои ограничения, далее рассматриваются перспективные альтернативы, позволяющие более объективно выявлять и исключать хайпы на основе данных. Например, исследователи рассматривают возможности использования комплексной системы метрик, отражающих интерес к новым технологиям в различных источниках, для оценки местоположения технологий на хайп-кривой Gartner [13]. Возможность автоматизированного моделирования хайп-цикла также подтверждается в работе [14], где авторы при помощи методов машинного обучения проводят анализ текстовых данных (научных статей и патентов) для определения динамики местоположения на кривой Gartner такой технологии, как квантовый компьютер.

Наряду с описанными рассматриваются и альтернативные методики, например, структурированный качественный подход, основанный на применении критического реализма, для определения факторов, которые действительно влияют на успешное внедрение технологий. Для выявления и разграничения обоснованных прогнозов и чрезмерно оптимистичных ожиданий предлагается фокусироваться на оценке реальных, измеримых эффектов, которые могут последовать за внедрением определенной технологии: эффектах автоматизации, информационных (улучшение координации и принятия решений), а также трансформационных эффектах (трансформация рынков). Еще один пример использования качественного анализа для выявления и отделения хайпов от научно-обоснованной информации можно найти в методологии *Grounded Theory* (GT). Она применяется для систематического анализа литературы, субъективность которого сводится к минимуму, по блокчейн-технологиям. Обширный корпус литературы (включающий как научные, так и научно-популярные материалы) кодируется и категоризируется, после чего содержимое полученных категорий анализируется (в том числе на предмет обоснованности) и интерпретируется, позволяя выявлять хайпы в массиве текстовых данных.

Известна и смешанная методология, сочетающая применение модели ARIMA (интегрированной модели авторегрессии – скользящего среднего) и семантических сетей для выявления зарождающегося хайпа в текстовых медиа (блогах и газетах). Успешное применение подобного подхода говорит о возможности идентификации аномальных всплесков интереса в данных, в том числе и в области исследования технологий. Так, в работе [15] исследуются поведение розничных инвесторов на российском фондовом рынке, а также их склонность к иррациональным действиям под влиянием социальных сетей и хайпов вокруг определенных ценных бумаг. Авторы вводят понятие индикатора хайпа, который интегрирует рассчитываемые при помощи методов машинного обучения метрики внимания (частота упоминаний компании) и настроения (позитивная или негативная тональность сообщений в отношении компании) на основе текстовой информации из социальных сетей. По результатам исследования индикатор хайпа продемонстрировал способность выявлять необоснованные ожидания инвесторов и способствовать построению прибыльных инвестиционных стратегий. Таким образом, подобный подход позволил не только идентифицировать компании, вокруг которых формируется необоснованный ажиотаж на фондовом рынке, но и сигнализировать о возможности построения аналогичных индикаторов хайпа для технологической сферы с целью эффективного автоматизированного выявления ажиотажа вокруг технологий.

Таким образом, хайпы, зачастую оказывающиеся продуктом преувеличенных ожиданий (влиять на которые также могут манипуляции общественным сознанием при помощи и технологий ИИ), способны искажать результаты технологической аналитики, создавая ложные представления о возможностях и перспективах развития технологий. Это, в свою очередь, может нести риски для процесса принятия решений. В литературе предлагаются различные инструменты для выявления хайпов в данных с помощью как качественного анализа, так и автоматизированных количественных методологий. Отделение хайпов от обоснованной информации в условиях роста объемов данных является нелинейной в решении задачей и требует комплексных подходов к выявлению чрезмерно завышенных ожиданий. Сочетание методов анализа больших данных, реализуемых в системе iFORA, и экспертного знания помогает выявлять подозрительные явления по временным всплескам ажиотажа, разграничивать их от устойчивых трендов, а также подтверждать или опровергать соответствующие гипотезы в задачах стратегического планирования.

Методика анализа больших текстовых данных для автоматизированного выявления хайповых технологий

Подход к автоматизированному выявлению хайпов на основе данных реализован в Институте статистических исследований и экономики знаний НИУ ВШЭ с использованием собственной разработки – системы интеллектуального анализа больших данных iFORA. Это уникальная коллекция, включающая более 800 млн документов из различных источников: научные публикации, патенты, документы рыночной аналитики и профессиональных СМИ, гранты и заявки на них, аналитические и стратегические документы, вакансии, образовательные программы и др. Ключевые типы документов представлены на русском, английском и китайском языках. Семантический анализ данных основан на технологиях глубокой обработки естественного языка, интегрирующих передовые методы машинного обучения и инструменты ИИ. Система основана на модульном подходе и позволяет комбинировать специализированные блоки в зависимости от решаемых аналитических задач для поддержки принятия управленческих решений в сфере науки, технологий и инноваций. Одним из таких блоков является хайп-анализ, который включает извлечение из больших массивов текстов тематик, обозначающих в большинстве случаев технологии или технологические направления, а также выявление среди них технологий, подверженных рискам завышенных ожиданий, то есть хайпов.

Извлечение технологических тематик.

Надежное извлечение из больших массивов неструктурированных текстов терминов, обозначающих технологии и технологические направления, является результатом разработки специальных языковых (эмбеддинговых) моделей, пригодных для выделения из терминологических полей высокоспециализированных терминов (обозначающих не только технологии, но в общем случае и произвольные заданные артефакты). Извлечение терминов предполагает работу с предварительно структурированными текстовыми документами, для чего каждый документ разделяется на отдельные предложения, слова и словосочетания с различными лингвистическими характеристиками. Сами документы отбираются из БД iFORA с использованием специально заданного поискового условия (набора понятий, в достаточной мере описывающих анализируемую тематическую область) и алгоритмов «умного» семантического поиска, учитывающего контекст употребления релевантной лексики в документах. Синтаксический анализ связей между словами в предложениях таких документов позволяет выявить зависимости слов (где одно слово является управляющим, а другое зависимым) и тип их связи, основанный на контексте. Далее для отдельных слов и словосочетаний вычисляются универсальные числовые характеристики в целях перехода от лингвистического аппарата к статистическому и наоборот, что позволяет отделить термины, содержащие информацию об объекте изучения, от лингвистического шума. Наиболее часто используемыми и эффективными показателями для определения высокоинформативных терминов являются частота слов, C-value, T-score и т.д. Также о высоком качестве результатов свидетельствует такой показатель, как поточечная взаимная информация. Это относительно простая логарифмическая линейная функция, типичная для использования в нейронных сетях с целью создания высококачественных векторных представлений терминов на больших корпусах данных.

Выявление хайпов.

К извлеченным из текстов технологическим тематикам далее применяется хайп-анализ, который позволяет разделять исследуемые тематики на две категории: обоснованные тренды/технологии и хайпы, не имеющие под собой значимой научно-технической базы исследований и разработок. Другими словами, хайп-анализ служит инструментом для разграничения устойчивых трендов и кратковременных всплесков необоснованного общественного интереса к определенным темам, которые становятся излишне популярными в медиапространстве. Результаты хайп-анализа визуализируются в виде хайп-карты (рис. 1).

Хайп-карта представляет собой диаграмму рассеяния тематик, расположенных в пространстве двух осей по рассчитанным для них значениям семантических показателей значимости в двух источниках – научных публика-

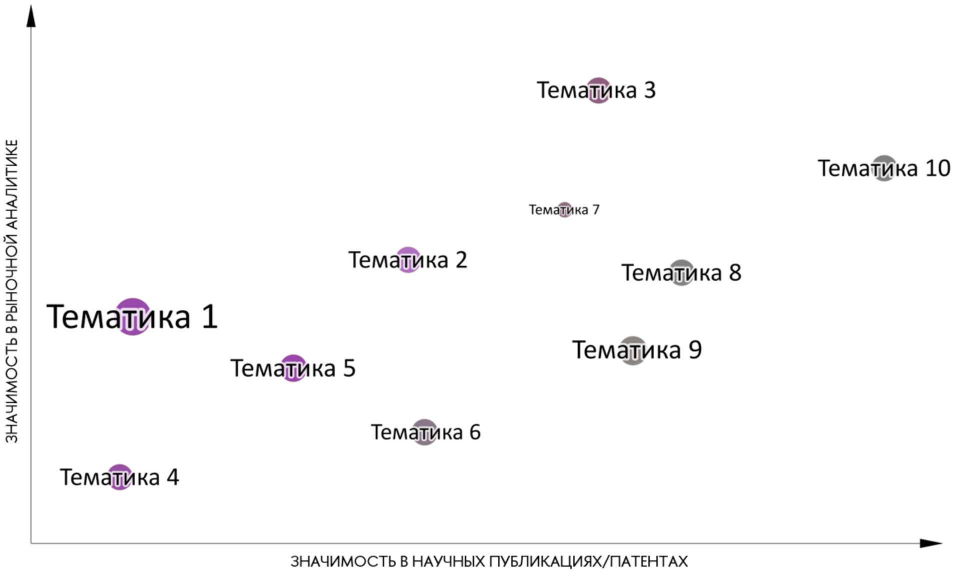

Рис. 1. Пример хайп-карты

Fig. 1. Example of a hype map

циях/патентах и документах рыночной аналитики и профессиональных СМИ. Значимость ($freq_i$) отражает интенсивность упоминаемости тематики в документах источника за выбранный период с учетом размера анализируемого корпуса документов:

$$freg_i = \frac{n_i}{N_i}, \quad (1)$$

где n_i – встречаемость термина в i-й год; N_i – размер корпуса документов в i-й год.

Отдельные тематики отражаются на хайп-карте круговыми значками. Высокий показатель значимости тематики указывает на ее сильную представленность в массиве документов. По оси абсцисс на хайп-карте представляются относительные показатели значимости тематик в научной или технологической повестке (рассчитываются на основе анализа научных публикаций или патентов соответственно, в отдельных случаях – по совокупному значению научных публикаций и патентов), по оси ординат – в медиаповестке (на основе анализа документов рыночной аналитики и профессиональных СМИ). За счет такого сравнения одни тематики могут быть определены как устойчивые тренды/технологии, а другие – как потенциальные хайпы.

При интерпретации данной визуализации к хайпам могут быть отнесены тематики с относительно высокой значимостью в медиапространстве и низкой в научных публикациях/патентах.

Такие тематики характеризуются высоким показателем хайповости (H), рассчитываемым как отношение значимости в двух анализируемых источниках (как правило, рассматривается отношение $freg_m$ – значимости тематики в документах рыночной аналитики и профессиональных СМИ к $freg_s$ – значимости в научных публикациях):

$$H = \frac{freg_m}{freg_s}. \quad (2)$$

Чем выше данный показатель, тем с большей вероятностью ту или иную тематику можно отнести к категории хайпов. В пространстве двух осей такие тематики обычно расположены в левой верхней части хайп-карты. Размер и цвет фигурных значков могут быть одинаковыми, а могут задаваться в системе экспертно и отражать различные семантические показатели для углубленной аналитики. Таким образом, хайп-карты наглядно визуализируют отличия в уровне значимости тематик в разных источниках данных. Показатель хайповости и расположение тематик на визуализации предоставляют исследователям информацию об их относительно большей популярности в медиаповестке по сравнению с научной/технологической. Такое сравнение по типам источников позволяет выявлять гипотезы о тематиках-хайпах, вызывающих необоснованно высокий интерес. При этом целесообразно проведение дополнительного анализа хайпов для формирования более надежных выводов о причинах и предполагаемых последствиях их возникновения.

Подход к автоматизированному выявлению хайпов с использованием системы iFORA апробирован на массиве текстовых данных по направлениям исследований и разработок в сфере цифровых технологий. В рамках проведенного исследования были использованы два типа источников информации на английском языке за 2017–2022 гг., в т.ч. база научных публикаций Microsoft Academic Graph и агрегированная база зарубежных отраслевых профессиональных СМИ и рыночной аналитики.

Полученная хайп-карта представлена на рисунке 2.

Направления исследований и разработок обозначены на визуализации числами от 1 до 50 и отражены в таблице. Для каждой тематики в ней также представлен рассчитанный показатель хайповости, таблица отсортирована по убыванию данного показателя.

Кроме табличных данных о степени хайповости исследованных тематик (и связанном с нею местоположении тематики на карте), данный показатель иллюстрируется цветами круговых значков, обозначающих тематики: градиентная цветовая маркировка круговых значков указывает на степень хайповости тематик, а также позволяет визуально сравнить их между собой. Круговые значки, обозначающие тематики с высоким индексом хайповости, тяготеют к фиолетовому цвету, с низким – к серому. Размер фигурных значков отражает показатель векторной центральности тематик (усредненной семантической близости тематики со всеми другими тематиками, включенными в анализ).

По результатам проведенного анализа определены направления исследований и разработок в сфере цифровых технологий, относящиеся к устойчивым трендам (среди которых имитационное моделирование (№ 20), геоинформационные системы (№ 39), обучение нейросетей (№ 2), фотоника (№ 38), технологии человеко-машинного взаимодействия (№ 46) и др.), а также выявлены гипотезы о хайповых направлениях. К топ-6 тематикам с наибольшим показателем хайповости отнесены следующие: приложения метавселенных (№ 30), цифровые аватары (№ 42), бескодовая разработка программного обеспечения (low-code / no-code, № 33), цифровые финансовые активы (№ 32), речевая аналитика (№ 6), децентрализованные финансовые сервисы (DeFi) (№ 23).

Высокий показатель хайповости указанных направлений исследований и разработок может быть обусловлен рядом факторов, связанных с недостаточной зрелостью потребительских рынков. Прежде всего к ним относятся отсутствие или низкая степень сформированности конечных рыночных продуктов, затрудняющие массовое внедрение технологий. Так, хотя отдельные решения в области метавселенных существуют, они крайне ограничены и не достигают массового пользователя, а сама технология остается на стадии концептуального развития. Это также касается технологии цифровых аватаров, которая, несмотря на потенциальные преимущества, в том числе в сферах образования и рекламы, пока не предлагает продукты,

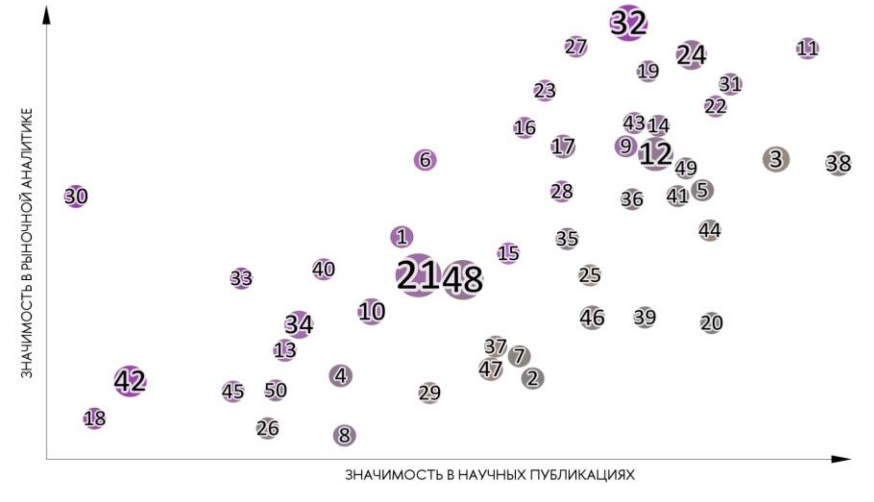

Рис. 2. Хайп-карта по 50 направлениям исследований и разработок в сфере цифровых технологий

Fig. 2. Hype map for 50 research and development areas in the field of digital technologies

Обозначение направлений исследований и разработок в сфере цифровых технологий, представленных на хайп-карте

Noting research and development areas in the field of digital technologies on the hype map

Лейбл кругового значка на хайп-карте – Соответствующее направление исследований и разработок	Показатель хайповости	Лейбл кругового значка на хайп-карте – Соответствующее направление исследований и разработок	Показатель хайповости
30 – Приложения метавселенных	888,615	43 – Биометрические технологии	3,704
42 – Цифровые аватары	367,054	22 – Облачные технологии	3,340
33 – Бескодовая разработка программного обеспечения (low-code / no-code)	283,994	14 – Сенсорные системы	2,995
32 – Цифровые финансовые активы	94,551	11 – Интернет вещей	2,968
06 – Речевая аналитика	70,846	15 – Промышленная робототехника	2,965
23 – Децентрализованные финансовые сервисы (DeFi)	28,375	12 – Умный город	2,841
01 – Интеллектуальный анализ данных	25,276	04 – Системы автоматического принятия решений	2,638
21 – Кибербезопасность	22,544	49 – Аддитивные технологии	2,110
18 – Беспилотная агротехника	18,718	35 – Технологии представления и доставки образовательного материала	1,982
16 – Беспилотный наземный транспорт (автомобили и общественный транспорт)	17,604	26 – Системы хранения разнородных больших данных	1,919
40 – Спутниковые системы связи	16,577	36 – Квантовые вычисления	1,839
44 – Интеллектуальные голосовые помощники	11,925	41 – Системы позиционирования и навигации	1,188
19 – Высокоскоростные технологии беспроводной передачи данных на малых дистанциях (WiFi, Bluetooth и др.)	11,382	05 – Анализ и обработка естественного языка	1,012
24 – Технологии распределенного реестра	9,120	08 – Автоматизированные системы медицинской диагностики	0,998
34 – Технологии управления жизненным циклом продукта	9,026	25 – Распределенные вычисления	0,884
17 – Технологии беспроводной передачи данных (5G, 6G и др.)	8,719	29 – Преобразование и кодирование данных	0,781
13 – Коллаборативные технологии	7,249	37 – Квантовые коммуникации	0,752
45 – Автоматическая идентификация и сбор данных	6,491	03 – Компьютерное зрение	0,734
21 – Цифровые двойники физических процессов	5,317	47 – Технологии управления образованием	0,589
10 – Когнитивные вычисления	5,110	07 – Системы идентификации и анализа звукового контента	0,465
28 – Системы аутентификации пользователей	4,396	46 – Технологии человеко-машинного взаимодействия	0,444
09 – Телемедицина	4,314	38 – Фотоника	0,354
50 – Адаптивное производство	4,204	02 – Обучение нейросетей	0,326
48 – Цифровые фабрики	3,914	39 – Геоинформационные системы	0,255
31 – Технологии виртуальной и дополненной реальности	3,858	20 – Имитационное моделирование	0,115

доступные для массового использования. В условиях высокой хайповости технологий компании могут делать значительные инвестиции в их развитие, опираясь на прогнозы о быстром возврате вложений. Тем не менее отсутствие реальных прикладных решений и их практической ценности может привести к неэффективности таких инвестиций и финансовым потерям. Кроме того, дефицит знаний у пользователей о возможностях и функциональности новых технологий является значительным препятствием для их распространения. Так, например, технологии бескодовой разработки могут позволить специалистам без профессиональных технических навыков создавать IT-продукты, но недостаток осведомленности о возможностях таких инструментов может препятствовать их активному внедрению в бизнес-процессы. В целях повышения конкурентоспособности компании могут интегрировать данные технологии в свои стратегии развития. Однако, если технологии окажутся временным хайпом, а не устойчивым трендом, возникнет необходимость пересмотра стратегий, что приведет к значительным финансовым потерям. Также следует отметить отсутствие четких регуляторных рамок в отношении хайповых технологий, что создает дополнительные барьеры для формирования полноценного рынка. Например, децентрализованные финансовые сервисы, открывающие доступ к широкому спектру финансовых инструментов, ограничены в использовании из-за неопределенности в отношении правовых аспектов их внедрения. Наряду с децентрализованными финансами цифровые финансовые активы также характеризуются невысокой практической применимостью, что может быть обусловлено недостаточной интеграцией в существующие финансовые системы и неопределенностью правовых аспектов их использования. Таким образом, ограниченность прикладных решений, дефицит знаний у пользователей и отсутствие четких регуляторных рамок могут способствовать высокой хайповости технологий. Данные факторы указывают на отсутствие или низкую зрелость массовых рынков, а активное медиаосвещение технологий может способствовать инвестициям в технологии, которые не будут реализованы в виде коммерческих продуктов.

Заключение

В настоящей работе рассмотрены возможности автоматизированного выявления хайпов среди трендов и технологий на основе анализа больших текстовых данных. Апробация предложенного подхода реализована в рамках семантического картирования направлений исследований и разработок на примере сферы цифровых технологий. С использованием системы iFORA проанализированы англоязычные научные публикации и документы отраслевой рыночной аналитики и профессиональных СМИ за 2017–2022 гг.

Результаты проведенного анализа позволили выделить как крупные системообразующие направления цифровых технологий, широко представленные в зарубежных научных публикациях, так и гипотезы о технологиях-хайпах, характеризующихся всплесками интереса в медиа. Так, к числу наиболее хайповых отнесены направления исследований и разработок, связанные с развитием метавселенных, цифровых аватаров и бескодовой разработки программного обеспечения. Направления, связанные с метавселенными и цифровыми аватарами, несмотря на значительное внимание со стороны общественности, к концу 2024 г. не демонстрируют устойчивость и масштабируемость, необходимые для трансформации в долгосрочные тренды. В то же время определение технологий бескодовой разработки программного обеспечения как хайпа в 2022 г. сегодня может быть интерпретировано как отражение обоснованно растущего общественного интереса, связанного с закономерным этапом развития технологии. Активная медийная поддержка данного направления подчеркивала перспективность и способствовала его становлению в качестве устойчивого тренда.

Таким образом, анализ больших текстовых данных можно рассматривать в качестве эффективного инструмента для частично автоматизированного выявления технологических хайпов. Для минимизации рисков, связанных с определением приоритетных направлений развития на основе их текстового следа, рекомендуется проводить углубленный анализ выявляемых хайповых трендов и технологий, учитывая их развитие в динамике. Такой подход позволит более точно различать краткосрочные всплески интереса от устойчивых тенденций, своевременно реагировать на изменения технологических рынков и принимать обоснованные решения относительно инвестиций и стратегического планирования.

Список литературы

1. Logue D., Grimes M. Living up to the hype: How new ventures manage the resource and liability of future-oriented visions within the nascent market of impact investing. AMJ, 2022, vol. 65, no. 3, pp. 1055–1082. doi: 10.5465/amj.2020.1583.

2. Intemann K. Understanding the problem of "hype": Exaggeration, values, and trust in science. Canadian J. of Philosophy, 2022, vol. 52, no. 3, pp. 279–294. doi: 10.1017/can.2020.45.

3. Yap J., McLellan B. A historical analysis of hydrogen economy research, development, and expectations, 1972 to 2020. Environments, 2023, vol. 10, no. 1, art. 11. doi: 10.3390/environments10010011.

4. White G.R.T., Samuel A. Programmatic advertising: Forewarning and avoiding hype-cycle failure. TFSC, 2019, vol. 144, pp. 157–168. doi: 10.1016/j.techfore.2019.03.020.

5. Shi Y., Herniman J. The role of expectation in innovation evolution: Exploring hype cycles. Technovation, 2023, vol. 119, no. 3, art. 102459. doi: 10.1016/j.technovation.2022.102459.

6. Chen X., Han T. Disruptive technology forecasting based on gartner hype cycle. Proc. TEMSCON, 2019, pp. 1–6. doi: 10.1109/TEMSCON.2019.8813649.

7. Kreps S., McCain R.M., Brundage M. All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. SSRN Electronic J., 2022, vol. 9, no. 1, pp. 104–117. URL: https://ssrn.com/abstract=3525002 (дата обращения: 23.06.2024). doi: 10.2139/ssrn.3525002.

8. Shoaib M.R., Wang Z., Ahvanooey M.T., Zhao J. Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models. Proc. ICCA, 2023, pp. 1–7. doi: 10.1109/ICCA59364.2023.10401723.

9. Nnamdi N., Oniyinde O.A., Abegunde B. An appraisal of the implications of deep fakes: The need for urgent international legislations. AJLG, 2023, vol. 8, no. 1, pp. 43–70. doi: 10.47672/ajlg.1540.

10. Korshunov P., Marcel S. Vulnerability assessment and detection of deepfake videos. Proc. ICB, 2019, pp. 1–6. doi: 10.1109/ICB45273.2019.8987375.

11. Burnaev E., Mironov E., Shpilman A., Mironenko M., Katalevsky D. Practical AI cases for solving ESG challenges. Sustainability, 2023, vol. 15, no. 17, art. 12731. doi: 10.3390/su151712731.

12. Ikram S.T., Priya V., Chambial S., Sood D., Arulkumar V. A performance enhancement of deepfake video detection through the use of a hybrid CNN Deep learning model. IJECES, 2023, vol. 14, no. 2, pp. 169–178. doi: 10.32985/ijeces.14.2.6.

13. Carbonell J., Sánchez-Esguevillas A., Carro B. Easing the assessment of emerging technologies in technology observatories. Findings about patterns of dissemination of emerging technologies on the internet. Technology Analysis & Strategic Management, 2018, vol. 30, no. 1, pp. 113–129. doi: 10.1080/09537325.2017.1337886.

14. Kondo Y., Asatani K., Sakata I. Extending hype cycle prediction by applying graph network analysis. Proc. PICMET, 2022, pp. 1–9. doi: 10.23919/PICMET53225.2022.9882551.

15. Teplova T., Tomtosov A., Sokolova T. A retail investor in a cobweb of social networks. PLoS ONE, 2022, vol. 17, no. 12, art. e0276924. doi: 10.1371/journal.pone.0276924.

Software & Systems doi: 10.15827/0236-235X.149.017-026 2025, 38(1), pp. 17–26

Automated identification of hype technologies: Semantic analysis

Irina V. Loginova [1]✉, Anna S. Piekalnits [1], Aleksandr V. Sokolov [1]

[1] National Research University Higher School of Economics, Moscow,101000, Russian Federation

For citation
Loginova, I.V., Piekalnits, A.S., Sokolov, A.V. (2025) 'Automated identification of hype technologies: Semantic analysis', *Software & Systems*, 38(1), pp. 17–26 (in Russ.). doi: 10.15827/0236-235X.149.017-026
Article info
Received: 12.07.2024 After revision: 23.07.2024 Accepted: 30.07.2024

Abstract. The research focuses on inflated public expectations of new technologies, or hypes. The paper presents the results of the development and testing an automated methodology for the automated identification of hypes among technological topics based on their textual trace in the digital technology field. The amount of new technological developments in the world is constantly growing, however their real potential for practical application varies greatly. Therefore, it is important to understand reliable factors to distinguish trends from hypes. Typically, industry and technology experts suggest that the possible signs of hypes include the following ones: absence of a stable business model, an unformed or obviously limited consumer market, and a large number of more effective alternatives. Identifying hypes in the technology agenda remains a difficult analytical task. This is due to the terminological inconsistency, the expert nature of the task, insufficiently developed methodological approaches, and the lack of specific technical tools. The method described in this paper involves the extraction of terms referring to technologies using natural language processing and computational linguistics techniques. These terms are extracted from several dozens of millions of different types of text documents, such as scientific publications, patents, and market analytics. The method also includes calculating an objective measure of each technology's "hype"

and constructing a visual map that illustrates the technology landscape that allows separating sustainable trends from potential hypes. Decision makers can use such hype maps in conjunction with other analytical results to identify priority development areas, analyze current and forecast future trends, and in risk management.
Keywords: semantic analysis, big data, trend analysis, technology analytics, hypes, risk assessment

References

1. Logue, D., Grimes, M. (2022) 'Living up to the hype: How new ventures manage the resource and liability of future-oriented visions within the nascent market of impact investing', *AMJ*, 65(3), pp. 1055–1082. doi: 10.5465/amj.2020.1583.
2. Intemann, K. (2022) 'Understanding the problem of "hype": Exaggeration, values, and trust in science', *Canadian J. of Philosophy*, 52(3), pp. 279–294. doi: 10.1017/can.2020.45.
3. Yap, J., McLellan, B. (2023) 'A historical analysis of hydrogen economy research, development, and expectations, 1972 to 2020', *Environments*, 10(1), art. 11. doi: 10.3390/environments10010011.
4. White, G.R.T., Samuel, A. (2019) 'Programmatic advertising: Forewarning and avoiding hype-cycle failure', *TFSC*, 144, pp. 157–168. doi: 10.1016/j.techfore.2019.03.020.
5. Shi, Y., Herniman, J. (2023) 'The role of expectation in innovation evolution: Exploring hype cycles', *Technovation*, 119(3), art. 102459. doi: 10.1016/j.technovation.2022.102459.
6. Chen, X., Han, T. (2019) 'Disruptive technology forecasting based on gartner hype cycle', *Proc. TEMSCON*, pp. 1–6. doi: 10.1109/TEMSCON.2019.8813649.
7. Kreps, S., McCain, R.M., Brundage, M. (2022) 'All the news that's fit to fabricate: AI-generated text as a tool of media misinformation', *SSRN Electronic J.*, 9(1), pp. 104–117, available at: https://ssrn.com/abstract=3525002 (accessed June 23, 2024). doi: 10.2139/ssrn.3525002.
8. Shoaib, M.R., Wang, Z., Ahvanooey, M.T., Zhao, J. (2023) 'Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models', *Proc. ICCA*, pp. 1–7. doi: 10.1109/ICCA59364.2023.10401723.
9. Nnamdi, N., Oniyinde, O.A., Abegunde, B. (2023) 'An appraisal of the implications of deep fakes: The need for urgent international legislations', *AJLG*, 8(1), pp. 43–70. doi: 10.47672/ajlg.1540.
10. Korshunov, P., Marcel, S. (2019) 'Vulnerability assessment and detection of deepfake videos', *Proc. ICB*, pp. 1–6. doi: 10.1109/ICB45273.2019.8987375.
11. Burnaev, E., Mironov, E., Shpilman, A., Mironenko, M., Katalevsky, D. (2023) 'Practical AI cases for solving ESG challenges', *Sustainability*, 15(17), art. 12731. doi: 10.3390/su151712731.
12. Ikram, S.T., Priya, V., Chambial, S., Sood, D., Arulkumar, V. (2023) 'A performance enhancement of deepfake video detection through the use of a hybrid CNN Deep learning model', *IJECES*, 14(2), pp. 169–178. doi: 10.32985/ijeces.14.2.6.
13. Carbonell, J., Sánchez-Esguevillas, A., Carro, B. (2018) 'Easing the assessment of emerging technologies in technology observatories. Findings about patterns of dissemination of emerging technologies on the internet', *Technology Analysis & Strategic Management*, 30(1), pp. 113–129. doi: 10.1080/09537325.2017.1337886.
14. Kondo, Y., Asatani, K., Sakata, I. (2022) 'Extending hype cycle prediction by applying graph network analysis', *Proc. PICMET*, pp. 1–9. doi: 10.23919/PICMET53225.2022.9882551.
15. Teplova, T., Tomtosov, A., Sokolova, T. (2022) 'A retail investor in a cobweb of social networks', *PLoS ONE*, 17(12), art. e0276924. doi: 10.1371/journal.pone.0276924.

Авторы
Логинова Ирина Владимировна [1],
заведующий отделом, iloginova@hse.ru
Пиекалнитс Анна Сергеевна [1],
ведущий эксперт, apiekalnits@hse.ru
Соколов Александр Васильевич [1], профессор,
заместитель директора, sokolov@hse.ru

[1] Институт статистических исследований
и экономики знаний НИУ
«Высшая школа экономики»,
г. Москва, 101000, Россия

Authors
Irina V. Loginova [1],
Head of Department, iloginova@hse.ru
Anna S. Piekalnits [1], Leading Expert,
apiekalnits@hse.ru
Aleksandr V. Sokolov [1], Professor,
Deputy Director, sokolov@hse.ru

[1] National Research University
Higher School of Economics,
Moscow, 101000,
Russian Federation