

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ТЕКСТОВ ГЕОЛОГИЧЕСКИХ ПУБЛИКАЦИЙ С ИСПОЛЬЗОВАНИЕМ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

М. И. Патук¹ [0000-0003-3036-2275], В. В. Наумова² [0000-0002-3001-1638]

^{1, 2}Государственный геологический музей им. В.И. Вернадского РАН, г. Москва,
Россия

¹patuk@mail.ru, ²naumova_new@mail.ru

Аннотация

Стремительный рост объема публикаций во всех областях геологических наук делает критически важным внедрение методов автоматизированной обработки научных текстов. Одним из наиболее перспективных инструментов для решения этой задачи выступают большие языковые модели на основе нейронных сетей. Огромный прорыв в области искусственного интеллекта за последние годы превратил такие модели в незаменимых помощников для исследователей. Наши работы по семантическому поиску публикаций с использованием дополнительно тренированных языковых моделей и нахождения меры близости геологических текстов показали хорошие результаты. Но используемые модели оказались неспособны выполнить глубокий анализ текстов. Сравнительный анализ современных архитектур позволил нам выделить модель DeepSeek R1, относящуюся к классу систем с расширенными возможностями логического вывода. Данный тип моделей демонстрирует принципиально новый уровень качества генерации. На базе выбранной модели разработан веб-сервис, предоставляющий уникальный функционал, осуществляющий сравнительный анализ до 5 научных статей стандартного объема; поддержку мульти язычных источников (ввод текстов на английском, китайском, русском и др. языках); формирование структурированных отчетов на русском языке с выделением ключевых тезисов, противоречий и паттернов. Проведено тестирование предложенного подхода для сравнительного анализа геологических публикаций. Тестирование показало результаты, вызывающие доверие.

Ключевые слова: искусственный интеллект, большие языковые модели, обработка естественного языка, анализ текстов, геология.

ВВЕДЕНИЕ

Стремительный рост объема публикаций во всех областях геологических наук является серьезным вызовом для исследователей, что заставляет уделять все больше времени изучению литературы. Проблема дополнительного инструмента, позволяющего автоматизировать процесс анализа периодических научных изданий, давно назрела. По нашему мнению, таким инструментом могут являться языковые модели и, особенно, большие языковые модели [1].

Наши работы в области анализа научных публикаций начинались с самых простых задач – бинарной классификации и реферирования текстов [2]. Были решены задачи классификации месторождений золота и железа по аннотациям статей и выделения ключевых слов из них. Эти задачи решались с помощью дополнительно тренированных языковых моделей относительно небольшого объема.

Следующей задачей, которая являлась более практически значимой, была задача семантического поиска [3]. Для ее решения была выполнена дополнительная тренировка другой языковой модели, которая была предназначена для нахождения семантической близости документов.

На основе введенного поискового запроса происходит поиск наиболее семантически близких публикаций в репозитории Государственного геологического музея (ГГМ) им. В. И. Вернадского РАН [4]. Поиск близости ведется по названиям и аннотациям статей.

С помощью этой же языковой модели была решена еще одна задача – поиск близости аннотаций двух статей геологической направленности. В качестве параметра близости рассчитывалась косинусная близость двух текстов. Но числовой параметр близости мало информативен, поэтому было решено найти более развернутую оценку схожести и различия двух текстов. Данная задача также была реализована, и мы получили текстовое описание сходства и различия двух текстов (рис. 1).

На рис. приведен фрагмент описания различий двух месторождений алмазов – трубки «Мир» [5] и трубки «Аргайл» [6] (Австралия). Обращают на себя внимание некорректные утверждения типа: «тип месторождения не указан в тексте», «текст не содержит деталей геологии», «в тексте не описываются конкретные характеристики». Данный недостаток является очевидным, т. к. мы использовали

для анализа аннотации статей, а не полные тексты. С другой стороны, мы были ограничены параметрами используемых языковых моделей, контекстное окно (количество текста для анализа) которых ограничено 512 токенами (около 200–250 слов русского языка, в английском немного более).

Различия:

1. Тип месторождения и геология:

- **Мир:** Четко указан тип - **кимберлитовая трубка**. Текст фокусируется на **изменениях геологического строения с глубиной** (увеличение объема рудного тела >1500 м, стабильность состава).
- **Аргайл:** **Тип месторождения не указан в тексте** (известно, что это **лампроитовая трубка**, что и было вызовом традициям). Текст **не содержит деталей геологии**, акцент на историю добычи и характеристики продукции.

2. Характер продукции (Алмазы):

- **Мир:** **В тексте не описываются конкретные характеристики** добываемых алмазов (размер, цвет, качество). Упор на объем руды и запасы алмазов в каратах.

Рис. 1. Текстовое описание различий аннотаций двух статей.

Стало очевидно, что для анализа полного текста статей нужно переходить на языковые модели другого класса – большие языковые модели. Модели предыдущего поколения, которые мы использовали, могли обрабатывать небольшие объемы текста и были ориентированы на узкий круг задач обработки естественного языка. Это было плюсом с точки зрения адаптации их под конкретные задачи пользователя. Можно было дообучать их на своих наборах данных, размещать на локальных серверах. Но, как мы показали выше, для них есть свои естественные ограничения.

АНАЛИЗ ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ

Большие языковые модели (LLM) стали очень бурно развиваться с начала 2020-х годов. Особенно большой резонанс вызвал выход в 2022 г. первой версии модели ChatGPT [7]. Данной модели и ее использованию в научной деятельности посвящено огромное количество публикаций. Большая часть исследователей в академических кругах обратила внимание в первую очередь на генеративные свойства этой модели.

ChatGPT может создавать тексты на заданные темы и неплохо владеет научной терминологией. Исследование 45000 статей, вышедших за 3 месяца после выхода ChatGPT, показало, что около 10% статей использовали его в той или иной степени [8]. Большинство авторов сходится во мнении, что ChatGPT может быть полезен в автоматизации рутинных задач, таких как: анализ текстов, перевод, классификация, обзор литературы [9]. Хотя по поводу обзоров литературы не все авторы разделяют оптимизм [10, 11]. Сравнение результатов ChatGPT с реальными исследованиями показало существенную неполноту искусственного интеллекта (ИИ) и ограниченность выборки материалов датой подготовки материалов для обучения модели. Особенно неприятным явлением стали так называемые «галлюцинации» [12], вымышленные ответы системы, не имеющие отношения к действительности.

Со временем появилось понимание того что эти модели, по крайней мере в их текущем состоянии, надо использовать как помощников в исследованиях [13]; выполнять с их помощью трудоемкие, но рутинные задачи. При этом от пользователя таких инструментов требуется глубокое понимание предметной области, чтобы направлять ИИ и проверять его выводы.

Вместе с тем не все исследователи признают первенство ChatGPT над другими моделями. Там, где требуется глубокое знание предметной области, дополнительно обученные модели предыдущего поколения, типа BERT, способны показывать лучшие результаты [14]. Авторы решали традиционную для обработки естественного языка задачу классификации кернов скважин по типам осадочных пород на основании текстового описания этих кернов. При этом была поставлена довольно сложная задача – кроме основного литологического типа выделять еще

два дополнительных. Их дополнительно обученная модель показала результаты классификации в среднем на 11% лучше, чем GPT-4.

МЕТОДИКА РАБОТЫ

В настоящее время существует большое количество различных больших языковых моделей [15]. Но, к сожалению, многие из них нам недоступны в силу различных причин. Мы выбрали для анализа три модели: T-lite-it-1.0-Q8_0-GGUF [16] от T-банка, GigaChat [17] от Сбера и китайскую модель DeepSeek [18]. Первую модель можно было запускать локально, две другие вызывались по API.

С учетом нашей задачи – сравнительного анализа геологических текстов – мы построили методику анализа работы указанных моделей.

Все современные большие языковые модели являются многоязычными и поддерживают перевод текстов. Были использованы несколько пар текстов описания различных месторождений на русском и английском языках. Указанные тексты подавались на вход моделям в режиме чат-бота с основным вопросом: описать сходство и различие полученных текстов. После получения ответа он анализировался на соответствие действительности. Кроме того, задавались различные вопросы с целью проверки качества запоминания каждого отдельного текста. Например, если в качестве сравниваемых текстов выступали описания месторождений золота и железа, мы просили модель указать, в каком из текстов содержится описание минерала «гематит» (по факту в описании месторождения железа) или описание минерала «арсенопирит» (по факту в описании месторождения золота).

По результатам этих тестов модель от Т-банка показала худшие результаты. Она плохо запоминала каждый отдельный текст и показывала большое количество «галлюцинаций», т. е. выдуманных фактов. Модель от Сбера оказалась немного лучше. Но безусловным лидером оказалась модель DeepSeek R1. Она относится к классу систем с расширенными возможностями логического вывода. Перед каждым выводом модель демонстрирует цепочку своих рассуждений по теме поставленного вопроса. Она всегда четко запоминала каждый отдельный текст, не путалась в содержании текстов и выдавала четкий структурированный ответ. Данная модель была использована для создания веб-сервиса по сравнению текстов геологических статей.

ВЕБ-СЕРВИС

Указанный веб-сервис вместе с другими сервисами анализа текстов расположен на портале GeologyScience.ru Государственного геологического музея им. В. И. Вернадского РАН. Он доступен через пункты меню «Сервисы» – «ИИ-анализ геологических текстов» или по отдельному адресу <http://83.149.241.22/> (рис. 2).

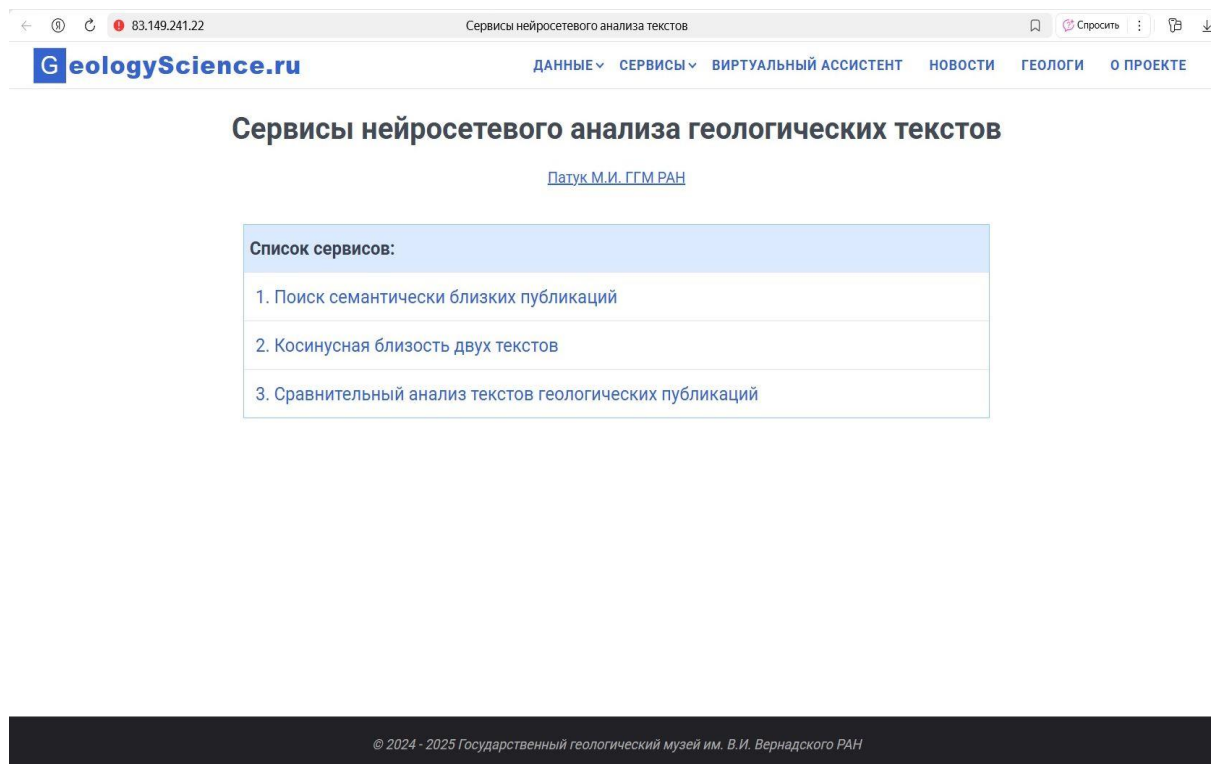


Рис. 2. Сервисы нейросетевого анализа текстов.

Сервис «Сравнительный анализ текстов геологических публикаций» принимает на вход статьи в формате PDF, которые должны сопровождаться основным вопросом типа: «Сравни эти два текста». Кроме того, можно задавать любое количество дополнительных вопросов: «В чем различие геологических описаний двух приведенных месторождений», «Какой возраст вмещающих пород первого месторождения» и т. п. Полученный PDF-файл преобразуется в текст и вместе с вопросом пользователя по API отправляется на сервер DeepSeek. Полученный ответ демонстрируется пользователю в окне сервиса.

Нами были повторно проанализированы две статьи по месторождениям алмазов – трубка «Мир» [5] и трубка «Аргайл» [6] (Австралия), которые не совсем удачно мы пытались анализировать с помощью своего предыдущего сервиса.

При этом вместо аннотаций анализу были подвергнуты полные тексты статей. Результат представлен ниже.

**Сравнение двух геологических текстов с помощью нейросетевой модели
DeepSeek R1**

**Сравнительный анализ алмазных месторождений: кимберлитовая трубка
«Мир» (Россия) и месторождение Аргайл (Австралия)**

1. Геологические особенности

Критерий	Трубка «Мир»	Аргайл
Тип породы	Кимберлитовая трубка	Лампроитовая трубка (первое крупное лампроитовое месторождение)
Возраст	Возраст кимберлитов: ~360 млн лет	Возраст лампроита: ~1.2 млрд лет; алмазы: ~1.58 млрд лет
Морфология	Удлиненно-овальная форма на глубине, расширение объема рудного тела с глубиной	Трубка в форме «головастика», наклонена на 30°, длина 2000 м
Геодинамическая обстановка	В пределах Якутской алмазодносной провинции, на краю Сибирского кратона	В зоне Холлс-Крик (окраина Кимберлийского кратона), тектонически активная зона

Ключевое отличие: Аргайл изменил парадигму геологии алмазов, доказав, что они могут формироваться в лампроитах, а не только в кимберлитах.

2. Объемы добычи и запасы

Критерий	Трубка «Мир»	Аргайл
Пиковая добыча	Нет точных данных, но за 40 лет карьером извлечено 350 млн м ³ породы	42.8 млн карат в 1994 г. (40% мировой добычи)
Средняя годовая добыча	На подземном руднике (после 2009 г.) — 1 млн т руды в год	~25–30 млн карат/год в 1990-х
Запасы	+60 млн карат после разведки глубоких горизонтов (общие запасы >60 млн карат)	61 млн т руды с содержанием 2.9 карат/т (2000 г.)
Содержание алмазов	~3–6 карат/т	3 карата/т (в 3–10 раз выше типичных кимберлитов)

3. Технологии добычи и переработки

Критерий	Трубка «Мир»	Аргайл
Способ добычи	Карьер (до 525 м) → подземный рудник «Мир-Глубокий» после 2017 г.	Открытая добыча (планировался переход на подземную к 2006 г.)
Методы разведки	Наклонно-направленное бурение, керновое опробование, ГИС	Бурение сетки скважин, тяжело средняя сепарация, рентгеновская сортировка
Переработка	Механическое дробление, гравитационное обогащение	Многоступенчатое дробление, тяжело средняя сепарация, рентгеновские сортеры

Инновации: Аргайл разработал уникальные рентгеновские сортеры для обработки мелких алмазов (до 200 камней/сек). В «Мире» акцент на разведку глубоких горизонтов с использованием телеметрии.

4. Маркетинг и продукция

Критерий	Трубка «Мир»	Аргайл
Основная продукция	Бесцветные и слабо-окрашенные алмазы	Мелкие коричневые/желтые (72%), редкие розовые (<1%)
Маркетинговые стратегии	Фокус на промышленное значение, увеличение запасов	Продвижение «шампанских» и «коньячных» алмазов (шкала C1–C7), аукционы розовых алмазов
Рынки сбыта	Внутренний (Россия), экспорт через АЛРОСА	Индия (огранка), глобальные аукционы (розовые алмазы)

Особенность: Аргайл создал новый сегмент рынка для цветных алмазов, тогда как «Мир» ориентирован на массовую добычу.

5. Экологические и социальные аспекты

Критерий	Трубка «Мир»	Аргайл
Экологические меры	Управление водными ресурсами после затопления 2017 г., «мокрая консервация»	Сертификация ISO 14001, рекультивация земель, создание озера из карьера
Социальное влияние	Основа алмазной промышленности России, создание рабочих мест в Якутии	Занятость 800+ человек, взаимодействие с аборигенами, вахтовый метод работы
Проблемы	Затопление шахты в 2017 г., необходимость перехода на подземную добычу	Высокая себестоимость при углублении карьера, удаленность месторождения

6. Проблемы и решения

- **«Мир»:**
 - *Проблема:* Катастрофическое затопление в 2017 г. остановило добычу.
 - *Решение:* Проект «Мир-Глубокий» для подземной разработки, интенсификация геологоразведки.

- **Аргайл:**
 - *Проблема:* Низкая рентабельность мелких алмазов.
 - *Решение:* Партнерство с Индией для огранки, создание премиального сегмента (розовые алмазы).
-

Выводы

- **Геология:** Аргайл расширил понимание алмазоносных пород, доказав роль лампроитов. «Мир» остается эталоном кимберлитовых месторождений.
- **Экономика:** Аргайл доминировал по объемам в 1990-х, но «Мир» обеспечивает долгосрочные запасы для России.
- **Инновации:** Оба месторождения используют современные технологии, но Аргайл преуспел в маркетинге, а «Мир» – в решении инженерных вызовов.
- **Устойчивость:** Аргайл лидирует в экологических практиках, «Мир» фокусируется на восстановлении после аварий.

Оба месторождения стали символами национальной гордости и драйверами развития своих регионов, демонстрируя разные пути адаптации к геологическим и рыночным вызовам.

ЗАКЛЮЧЕНИЕ

Как можно видеть, языковая модель DeepSeek R1 подробно по 6 основным пунктам выполнила анализ двух статей. Поскольку одна из статей была на английском языке, моделью был выполнен корректный перевод, что наглядно видно в приведенном анализе. Хотя следует отметить, что в процессе наших экспериментов модель иногда оставляла некоторые термины на английском языке или даже, очень редко, на китайском. Отметим, что мы не указывали, по каким пунктам следует выполнять сравнение. Они были установлены моделью на основе анализа текстов. Кроме того, были проверены все числовые данные: даты, геологический возраст, размеры карьеров, объемы добычи. Все цифры строго соответствуют тексту. Галлюцинации отсутствуют.

Размер контекстного окна модели DeepSeek равен 128 k токенов (60–65 тыс. слов русского языка). Это позволяет выполнять совместный анализ до 5 файлов

среднего размера (~10 страниц текста). Это будет предметом наших дальнейших исследований.

Применение модели DeepSeek для решения задач, с которых мы начинали свои работы по анализу текстов, бинарной классификации и извлечения ключевых слов, показало хорошие результаты, сопоставимые или превосходящие наши первые опыты. При этом эта модель не требовала дополнительно тренировки в предметной области для решения таких задач, поскольку перечисленные задачи были достаточно простыми.

Но обратим еще раз внимание на работу [14], в которой авторы показали превосходство дополнительно тренированной специализированной модели в задаче классификации осадочных пород над ChatGPT. И вслед за авторами этой работы повторим их вывод: если вы незнакомы с машинным обучением или у вас нет доступа к соответствующим техническим средствам, то, конечно, следует пользоваться такими инструментами, как ChatGPT или DeepSeek. Но специализированные модели в современных условиях могут дать гораздо лучший результат.

Кроме того, отметим специфику нашей задачи сравнительного анализа текстов для языковых моделей. Мы ограничиваем свободу генеративных возможностей модели текстами, предоставленными для анализа. Полученные выводы легко проверяются и в случае неверных ответов, как это имело место на этапе тестирования, легко заметить ошибки. Разумеется, для более глубокого анализа результатов работы больших языковых моделей по анализу геологических текстов требуется проверка ответов моделей со стороны специалистов в предметной области. Мы надеемся, что наша работа послужит импульсом для специалистов геологического профиля активнее применять подобные инструменты.

Наши тестовые результаты показывают, что современные большие языковые модели являются хорошим и доступным инструментом для анализа текстов научных публикаций. Но, как и любой инструмент, он требует дальнейшей проверки от заинтересованных специалистов.

Благодарности

Работа выполнена в рамках Государственной темы Государственного геологического музея им. В. И. Вернадского РАН (FMMR-2025-0001) Рег.

№1022061000107-4-1.5.1 «Естественно научные музеи в едином цифровом пространстве геологических знаний».

СПИСОК ЛИТЕРАТУРЫ

1. Большая языковая модель
URL: [https://ru.wikipedia.org/wiki/ Большая_языковая_модель](https://ru.wikipedia.org/wiki/Большая_языковая_модель) (дата обращения 01.10.2025)
2. Патук М.И., Наумова В.В. Методы искусственного интеллекта для научных исследований в геологии // Электронные библиотеки. 2023. Т. 26, №5. С. 673–696. <https://doi.org/10.26907/1562-5419-2023-26-5-673-696>
3. Патук М.И., Наумова В.В. Использование семантического поиска для выбора и ранжирования научных геологических публикаций // Электронные библиотеки. 2024. Т. 27, №5. С. 758–773.
<https://doi.org/10.26907/1562-5419-2024-27-5-758-773>
4. Патук М.И., Наумова В.В., Еременко В.С. Цифровой репозиторий "geologyscience.ru": открытый доступ к научным публикациям по геологии России // Электронные библиотеки. 2020. Т. 23, № 6. С. 1324–1338.
5. Килижеков О.К., Толстов А.В., Яхин Ш.М., Зырянов И.В. Месторождение алмазов кимберлитовой трубки Мир: основные этапы изучения, особенности и результаты разведки. Горная промышленность. 2025. № 1. С. 49–56.
<https://doi.org/10.30686/1609-9192-2025-1-49-56>
6. Shigley J., Chapman J., Ellison R. Discovery and Mining of the Argyle Diamond Deposit, Australia // Gems and Gemology. 2001. Vol. 37. P. 26–41.
<https://doi.org/10.5741/GEMS.37.1.26>
7. ChatGPT. <https://ru.wikipedia.org/wiki/ChatGPT> (дата обращения 01.10.2025)
8. Picazo-Sanchez P., Ortiz-Martin L. Analysing the impact of ChatGPT in research // Applied Intelligence. 2024. Vol. 54. P. 4172–4188.
<https://doi.org/10.1007/s10489-024-05298-0>
9. Islam I., Islam M.N. Exploring the opportunities and challenges of ChatGPT in academia // Discover Education. 2024. Vol. 3. Article no. 31.
<https://doi.org/10.1007/s44217-024-00114-w>

10. Faiza Farhat F., Sohail Sh. S., Dag Øivind Madsen D.Ø. How trustworthy is ChatGPT? The case of bibliometric analyses // Cogent Engineering. 2023. Vol. 10. Article no. 2222988. <https://doi.org/10.1080/23311916.2023.2222988>
 11. Зашихина И.М. Подготовка научной статьи: справится ли ChatGPT? // Высшее образование в России. 2023. Т. 32. № 8-9. С. 24–47. <https://doi.org/10.31992/0869-3617-2023-32-8-9-24-47>
 12. Галлюцинация (искусственный интеллект), URL: [https://ru.wikipedia.org/wiki/Галлюцинация_\(искусственный_интеллект\)](https://ru.wikipedia.org/wiki/Галлюцинация_(искусственный_интеллект)) (дата обращения 01.10.2025)
 13. Salvagno M., Taccone F.S., Gerli A.G. Can artificial intelligence help for scientific writing?// Critical Care. 2023. Vol. 27. Article no. 75. <https://doi.org/10.1186/s13054-023-04380-2>
 14. Ghorbanfekr H., Kerstens P.J., Dirix K. Classification of geological borehole descriptions using a domain adapted large language model // Applied Computing and Geosciences. 2025. Vol. 25. Article no. 100229.
 15. LLM Leaderboard, <https://artificialanalysis.ai/leaderboards/models> (дата обращения 01.10.2025)
 16. T-lite, https://huggingface.co/t-tech/T-lite-it-1.0-Q8_0-GGUF (дата обращения 01.10.2025)
 17. GigaChat. <https://giga.chat/> (дата обращения 01.10.2025)
 18. DeepSeek. <https://www.deepseek.com/en> (дата обращения 01.10.2025)
-

COMPARATIVE ANALYSIS OF GEOLOGICAL TEXTS USING LARGE LANGUAGE MODELS

Mikhail I. Patuk¹ [0000-0003-3036-2275], Vera V. Naumova² [0000-0002-3001-1638]

^{1, 2} *Vernadsky State Geological Museum of the Russian Academy of Sciences, Moscow, Russia*

¹patuk@mail.ru; ²Naumova_new@mail.ru

Abstract

The rapid increase in the volume of publications in various fields of geology makes it crucial to introduce methods for automated processing of scientific texts. Large language models based on neural networks represent one of the most promising approaches to solving this challenge. The recent breakthroughs in artificial intelligence have made such models indispensable tools for researchers. Our work on semantic search for publications using additionally trained language models and measuring the similarity between geological texts yielded good results. However, the models we used were unable to perform in-depth text analysis. A comparative analysis of modern architectures identified the DeepSeek R1 model as belonging to a class of systems with advanced logical inference abilities. This type of model represents a fundamentally new level of quality in text generation. Based on the chosen model, we have developed a web service that provides unique functionality for comparative analysis of up to 5 scientific articles. The service supports multilingual sources, allowing users to input text in English, Chinese, Russian, etc. It generates structured reports in Russian, highlighting key theses, contradictions, and patterns. The proposed approach has been tested on geological publications, and the results have been promising.

Keywords: artificial intelligence, large language models, natural language processing, texts analysis, geology.

REFERENCES

1. Large language model.
https://en.wikipedia.org/wiki/Large_language_model?ysclid=mg7ip9ev9d289421479
(date of access 01.10.2025)

2. *Patuk M.I., Naumova V.V.* Artificial Intelligence Methods for Scientific Research in Geology // *Russian Digital Libraries Journal*. 2023. Vol. 26, No. 5. P. 673–696. (In Russ.). <https://doi.org/10.26907/1562-5419-2023-26-5-673-696>
 3. *Patuk M.I., Naumova V.V.* Using Semantic Search to Select and Rank Geological Publications // *Automatic Documentation and Mathematical Linguistics*. 2024. Vol. 58, Suppl. 5. P. S294–S298. <https://doi.org/10.3103/S0005105525700372>
 4. *Patuk M.I., Naumova V.V., Eryomenko V.S.* Digital repository "geology-science.ru": open access to scientific publications on russian geology // *Russian Digital Library Journal*. 2020. Vol. 23, No. 6. P. 1324–1338 (in Russian).
 5. *Kilizhekov O.K., Tolstov A.V., Yakhin Sh.M., Zyryanov I.V.* Diamond deposit of the Mir kimberlite pipe: main research stages, specific features and results of exploration // *Russian Mining Industry*. 2025. No. 1. P. 49–56 (In Russ.). <https://doi.org/10.30686/1609-9192-2025-1-49-56>
 6. *Shigley J., Chapman J., Ellison R.* Discovery and Mining of the Argyle Diamond Deposit, Australia // *Gems and Gemology*. 2001. Vol. 37. P. 26–41. <https://doi.org/10.5741/GEMS.37.1.26>
 7. ChatGPT.
URL: <https://en.wikipedia.org/wiki/ChatGPT> (date of access 01.10.2025)
 8. *Picazo-Sanchez P., Ortiz-Martin L.* Analysing the impact of ChatGPT in research // *Applied Intelligence*. 2024. Vol. 54. P. 4172–4188. <https://doi.org/10.1007/s10489-024-05298-0>
 9. *Islam I., Islam M.N.* Exploring the opportunities and challenges of ChatGPT in academia // *Discover Education*. 2024. Vol. 3. Article no. 31. <https://doi.org/10.1007/s44217-024-00114-w>
 10. *Faiza Farhat F., Sohail Sh. S., Dag Øivind Madsen D.Ø.* How trustworthy is ChatGPT? The case of bibliometric analyses // *Cogent Engineering*. 2023. Vol. 10. Article no. 2222988. <https://doi.org/10.1080/23311916.2023.2222988>
 11. *Zashikhina I.M.* Scientific Article Writing: Will ChatGPT Help? Vysshee obrazovanie v Rossii // *Higher Education in Russia*. 2023. Vol. 32, no. 8. P. 24–47. <https://doi.org/10.31992/0869-3617-2023-32-8-9-24-47> (In Russ., abstract in Eng.)
 12. Hallucination (artificial intelligence). URL: [https://en.wikipedia.org/wiki/Hallucination_\(artificial_intelligence\)](https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence)) (date of access 01.10.2025)
-

13. *Salvagno M., Taccone F.S., Gerli A.G.* Can artificial intelligence help for scientific writing? // *Critical Care*. 2023. Vol. 27. Article no. 75.
<https://doi.org/10.1186/s13054-023-04380-2>
 14. *Hossein Ghorbanfekr, Pieter Jan Kerstens, Katrijn Dirix.* Classification of geological borehole descriptions using a domain adapted large language model // *Applied Computing and Geosciences*. 2025. Vol. 25. Article no. 100229.
 15. LLM Leaderboard.
<https://artificialanalysis.ai/leaderboards/models> (date of access 01.10.2025)
 16. T-lite. https://huggingface.co/t-tech/T-lite-it-1.0-Q8_0-GGUF (date of access 01.10.2025)
 17. GigaChat. <https://giga.chat/> (date of access 01.10.2025)
 18. DeepSeek. <https://www.deepseek.com/en> (date of access 01.10.2025)
-

СВЕДЕНИЯ ОБ АВТОРАХ



ПАТУК Михаил Иванович – к. г.-м. н., н. с., научный отдел Государственного геологического музея им. В.И. Вернадского РАН, Москва, Россия.

Michail I. PATUK – PhD, scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

email: patuk@mail.ru

ORCID: 0000-0003-3036-2275



НАУМОВА Вера Викторовна – д. г.-м. н., г. н. с., зав. Научным отделом Государственного геологического музея им. В.И. Вернадского РАН, Москва, Россия.

Vera V. NAUMOVA – Prof., head of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

email: naumova_new@mail.ru

ORCID: 0000-0002-3001-1638

Материал поступил в редакцию 30 августа 2025 года