Использование технологий искусственного интеллекта в работе с лицензионными ресурсами

Тенденции и практические кейсы

Генеративный ИИ: достоинства и недостатки

Достоинства и недостатки ГенИИ в библиотечном контексте обобщены в проекте документа ИФЛА “IFLA Toolkit on Libraries and Artificial Intelligence”. Среди главных недостатков (которых выделено 13) отмечена неточность и недостоверность информации (Inaccurate and misleading information). Галлюцинации (ИИ) – слово года в 2023 (по Кембриджу).

Основные причины генерации недостоверной информации
- ограничения в информации, на которой обучаются большие языковые модели;
- огромные объемы информации, на которых ***преимущественно статистическими методами*** вычисляются наиболее вероятные варианты текстов ответов.

Преодоление этих явлений породило развитие направления в генеративном ИИ, названного **Retrieval Augmented Generation (RAG)**. RAG применяется в сервисах поиска научной информации на базе двух основных принципов:
- запрос (промпт) посылается в проиндексированный массив ***научных публикаций***;
- ***ограниченное число наиболее релевантных документов*** становится основой для генерации ответов пользователю;
- в ответах содержатся ссылки на документы, из которых извлечена информация.

Примеры сервисов, применяющих RAG – Perplexity, Elicit, SciSpace и другие.

Иллюстрация: диалог с Perplexity

Do Elicit and SciSpace utilize RAG?

Auto — 8 sources

relevanceai
Implement Implicit RAG for Better AI Responses -...

chitika
RAG vs Finetuning vs Prompt Engineering: Key ...

deepset
Evaluating RAG Part I: How to Evaluate Document...

+5 sources

Answer

Both Elicit and SciSpace utilize the Retrieval Augmented Generation (RAG) technique to enhance their functionalities in research and academic contexts.

- **Elicit**: This AI research assistant employs RAG to automate and streamline research tasks such as summarizing papers, extracting data, and synthesizing findings. By integrating external information sources, Elicit improves the accuracy and relevance of the information it provides to users, making it a powerful tool for researchers [5] [8].
- **SciSpace**: Similarly, SciSpace leverages RAG to enhance its capabilities in academic search and literature review. By retrieving relevant information from external databases, it ensures that users receive up-to-date and contextually relevant responses, supporting better decision-making in research activities [8].

Both platforms exemplify how RAG can significantly improve the efficiency and effectiveness of research-related tasks by providing dynamic access to external knowledge bases.

Источники для тренировки ГенИИ

Для Perplexity, Elicit и SciSpace источник – преимущественно научные публикации, проиндексированные общедоступными сервисами (Semantic Scholar, Google Scholar, …). Это означает, что галлюцинации снижаются до минимума; некоторые авторы утверждают, что это понятие здесь вообще неприменимо, поскольку есть ссылки на документы, из которых взяты ответы: если ошибки есть, то они не порождаются, а извлекаются из этих документов.

Если рассматривать **научные ресурсы, лицензируемые библиотеками** для предоставления в доступ своим читателям, то в них источники тренировки ИИ еще более надежны, особенно для ресурсов, проводящих строгий отбор контента. Яркий пример – Scopus, которые тренировал Scopus AI на своем отборном контенте.

За последние два-три года несколько популярных и больших по объему ресурсов, которые лицензируются сотнями и тысячами библиотек по всему миру, внедрили специальные сервисы, использующие улучшенные технологии ГенИИ. Они соединяют два важных преимущества:

- надежная база источников для тренировки ИИ;
- Технология Retrieval Augmented Generation (RAG)

Примеры лицензионных ресурсов с встроенными сервисами ГенИИ

Primo Research Assistant- Clarivate, ProQuest)

Scopus AI - Elsevier

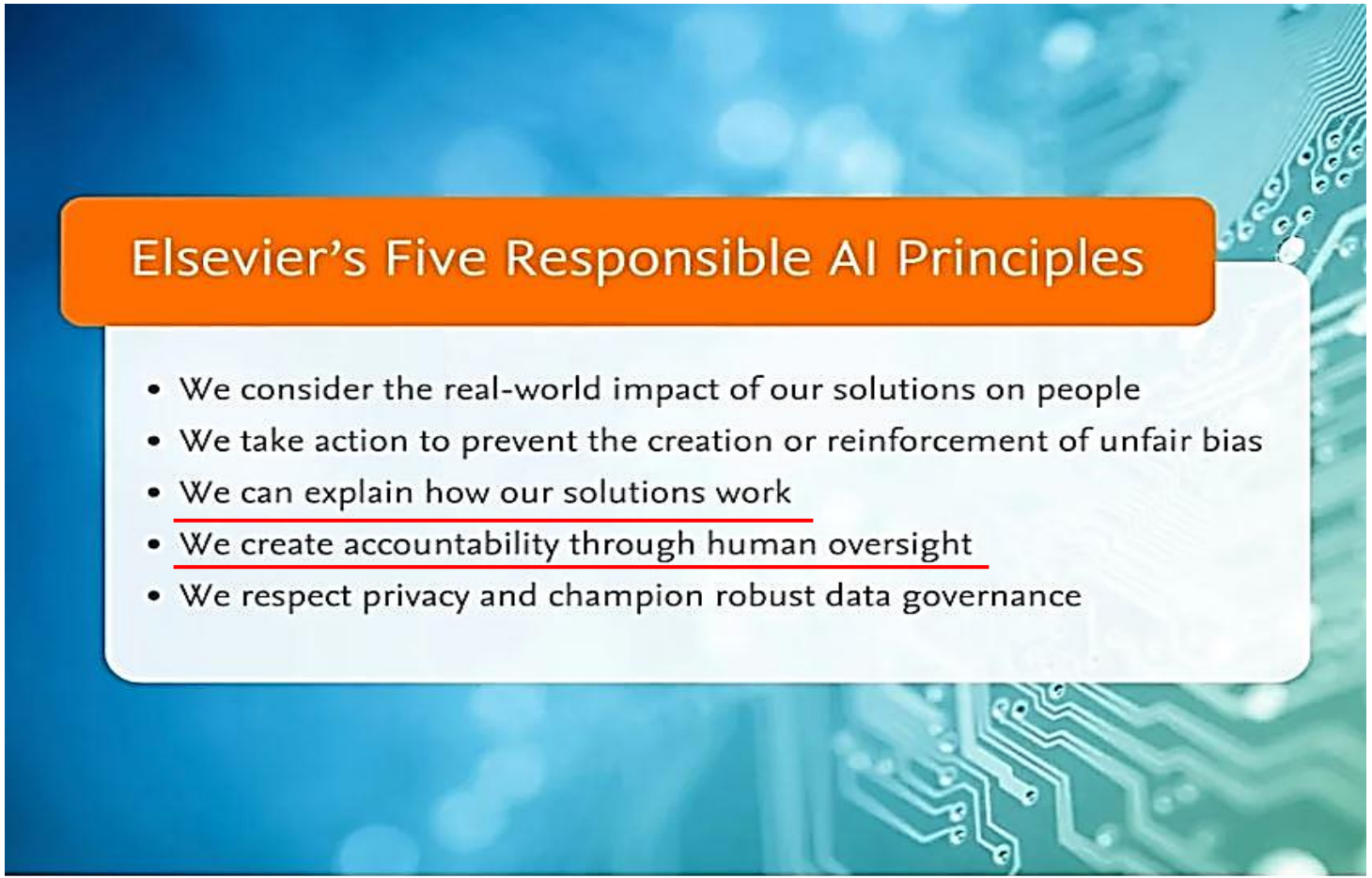
ScienceDirect AI (beta) - Elsevier

JSTOR Interactive Research Tool (beta) - ITHAKA

TDNet AI - TDNet

Все сервисы предоставляют четкие описания принципов их работы и контроля

Пять принципов ответственного применения ИИ издательства Elsevier



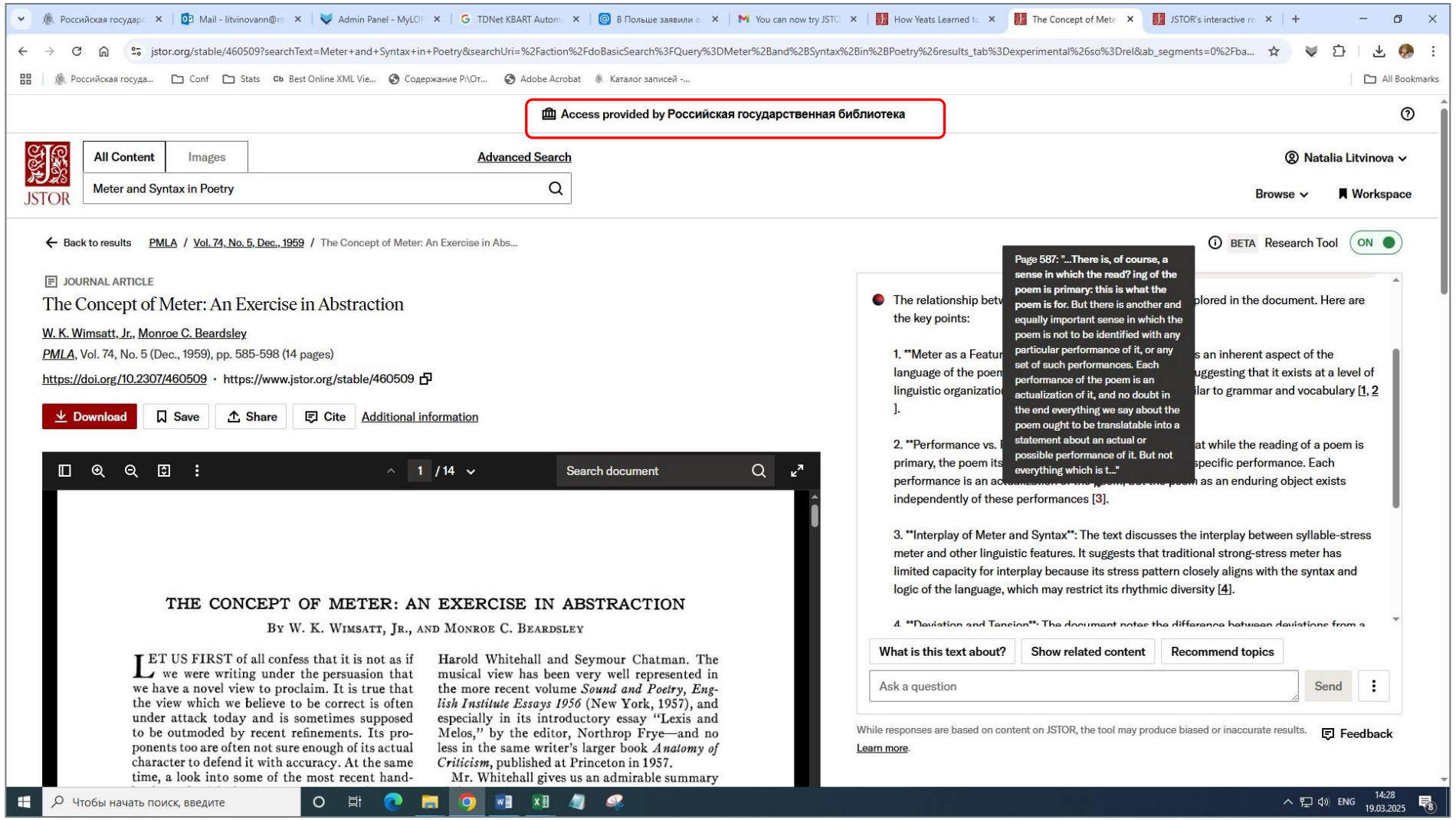
Общие характеристики сервисов с встроенными сервисами ГенИИ

- Разговорный (conversational) стиль взаимодействия: вопрос на естественном языке, уточнения в процессе диалога.
- Обязательная информация об источниках, на базе которых сформирован ответ: описания, ссылки.
- Обобщение (суммаризация) отдельных публикаций и/или всех публикаций, на основе которых сформирован ответ.

Кроме того, сервисы могут применять специфические опции на основе технологий ИИ. Например, Scopus AI строит карту понятий, используемых в публикациях.

Теперь задача наших библиотек – целенаправленно тестировать такие сервисы для использования в библиотечных процессах, прежде всего, в справочно-библиографическом обслуживании.

Иллюстрация: JSTOR – раскрытие основных тем документа



Практические кейсы (РГБ)

Кейс первый: использование ГенИИ для перевода заглавий и аннотаций книг издательства EBSCO Publishing, лицензированных РЦНИ для российских библиотек.

5167 книг научных зарубежных издательств были получены РГБ и в 2024 году загружены в электронный каталог и электронную библиотеку РГБ.
Загрузка в электронный каталог выполнена на основе метаданных, поставленных EBSCO. В результате основные поисковые поля (заглавие, предметные рубрики, аннотация) представлены только на языке оригинала (в основном английском). Нужно было обеспечить возможность поиска этих книг по запросам на русском языке. Кроме того, необходимо было сократить часть аннотаций большого объема (более 2000 знаков).

Эта задача решалась в два этапа:

1. 500 заглавий книг были переведены при помощи бесплатной версии DeepL Translate и оценены специалистом с лингвистическим образованием по трехбалльной системе (4 – хорошо, 3 – удовлетворительно, 2 – неудовлетворительно). В результате 78% получили оценку «хорошо», 20% - «удовлетворительно» и только 2% - неудовлетворительно. После этого по предложению комплекса информационных технологий было принято решение расширить эксперимент, сделав не только переводы заглавия, но и сокращение аннотаций до приемлемых объемов с последующим их переводом.
2. Результаты расширенного эксперимента также были оценены положительно (правда, на меньшем объеме); принято решение о реализации этого подхода силами комплекса информационных технологий. Идет загрузка отредактированных записей в ЭК РГБ.

Кейс второй: применение ГенИИ для оценки статистики использования ресурсов

В январе 2025 года введена в действие новая версия стандарта де факто для сбора и представления статистики использования электронных ресурсов – COUNTER 5.1. Среди нововведений этой версии – акцент на представление статистики использования ресурсов на уровне Item Report, то есть минимальных единиц потребления (загрузки) контента: статей, глав из книг и подобных. Эти данные представляются в отчетах IR (для любых единиц контента), IR_A1 (для статей) и IR_M1 (для мультимедийных объектов).

На этапе обсуждения новой версии предлагалось сделать эти отчеты обязательными, однако в финале они остались на уровне рекомендованных, но подкрепленных подчеркнутым вниманием к ним со стороны разработчиков и библиотечного сообщества.

Сейчас отчеты уровня Item Report предоставляют уже несколько десятков контент-провайдеров: Sage Publishing, Taylor&Francis, JSTOR, World Scientific Publishing и другие.

Интерес библиотек к отчетам уровня Item Report понятен: благодаря им можно получить более детальную информацию о востребованности ресурсов, прежде всего – о тематических предпочтениях пользователей, особенно в больших политематических коллекциях. Но для их получения необходимы специальные средства обработки данных. И здесь может помочь ГенИИ.

Статистика 2025 года пока доступна только за два месяца. Но некоторые издательства начали генерировать IR_A1 раньше. Данные этого отчета издателя Sage Publishing доступны за неполный 2024 год. Мы использовали их для эксперимента по индексированию названий статей индексами Универсальной десятичной классификации. Были взяты данные по первым 200 статьям, составившие в сумме около 500 выгрузок (25% годового использования).

Цель эксперимента - оценить качество индексирования по УДК заглавий статей с помощью GPT 4.0 с целью дальнейшего развития такого подхода к оценке статистики.

Фрагмент данных статистики Sage по статьям, проиндексированным по UDC (GPT 4)

Article title	UDC	Notation
The Workplace Commons: Towards Understanding Commoning within Work Relations	331.101	Labor relations and commons.
Beyond capitalist enclosure, commodification and alienation: Postcapitalist praxis as commons, social production and useful doing	330.342	Post-capitalist economic practices.
Development and Cross-Cultural Application of a Specific Instrument to Measure Entrepreneurial Intentions	658.11	Entrepreneurship and management tools.
Development of intelligent system of global bibliographic search	025.4:004	Bibliographic systems and artificial intelligence.
In plain sight: School librarian practices within infrastructures for learning	027.8:371.6	School libraries and educational infrastructures.
Labour as a Commons: The Example of Worker-Recuperated Companies	331.101	Labor relations; Worker cooperatives.
One nation, pulling apart: the basis of persistent poverty in the USA	330.567.2(73)	Poverty in the United States.
Pre-individual affects: Gilbert Simondon and the individuation of relation	1(091)	Philosophy and theories of individuation.
The multilingual children's library as physical and metaphorical 'space' within the community: Practical and emotional considerations	027.625	Multilingual children's libraries.
The Symptoms-Varices-Pathophysiology classification of pelvic venous disorders	616.14	Medical classification of venous disorders.

Первая десятка наиболее востребованных обобщенных тематик: 347 выгрузок

УДК	Статистика выгрузок	Нотация
316	69	Социология
331	55	Труд. Наука о труде. Экономика труда. Организация труда
911	43	Общая география. Отдельные отрасли географии. Ландшафтоведение
27	42	Христианство
330	30	Экономические науки в целом. Политическая экономия
616	29	Патология. Клиническая медицина
159	22	Психология
37	21	Народное образование. Воспитание. Обучение. Организация досуга
327	21	Международные отношения. Мировая политика. Внешняя политика
364	15	Общественные проблемы, порождающие необходимость оказания социальной помощи. Виды социальной помощи

Вывод: подход можно оценить как работоспособный; в развитие его необходимо разработать стандартные автоматизированные процедуры на базе отечественных сервисов ГенИИ.

Благодарю за внимание!