



Использование искусственного интеллекта в поиске научной информации

Федоров Андрей, Библиотека НИУ ВШЭ

Научно-практическая конференция «Трансформация вузовской библиотеки: стратегии и приоритеты», 5 апреля 2024 года

Большие языковые модели революционизируют науку и образование, обеспечивая мощные инструменты для анализа данных, создания новых знаний и улучшения образовательного и научного процесса



Обработка и анализ научных статей, поиск новых тенденций и направлений исследований, автоматизация процесса литературного обзора

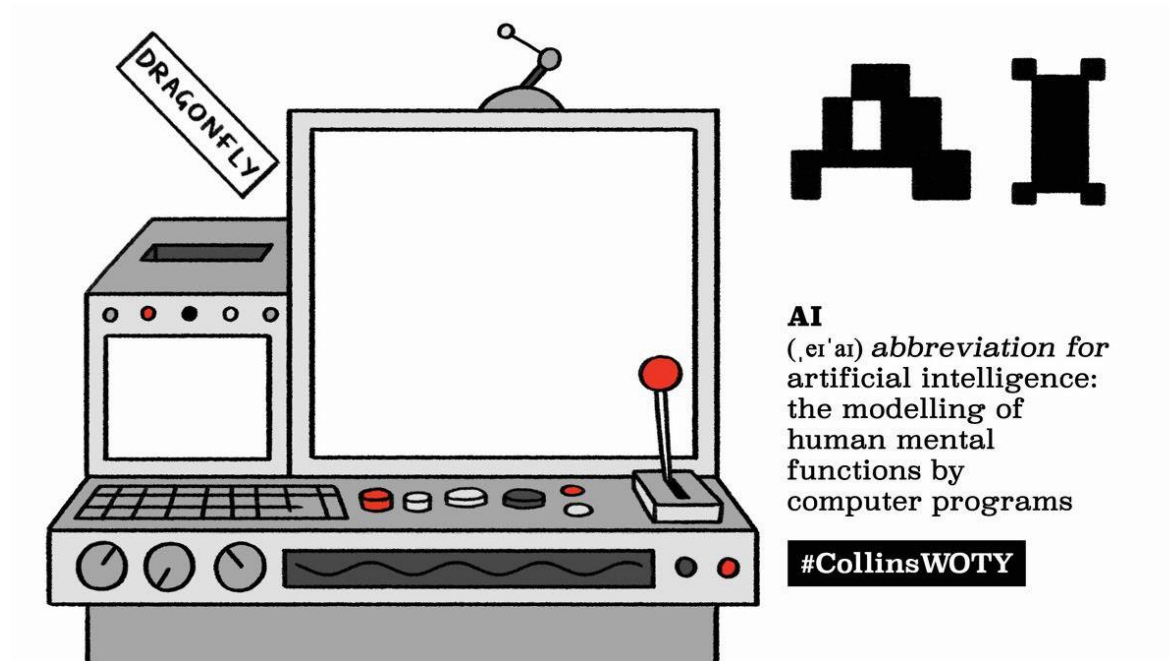


Персонализированное обучение, разработка учебных материалов, автоматизация оценки студентов и обратной связи

Определение понятия искусственного интеллекта

3

Искусственный интеллект – термин, описывающий моделирование умственных функций человека с помощью компьютерных программ*

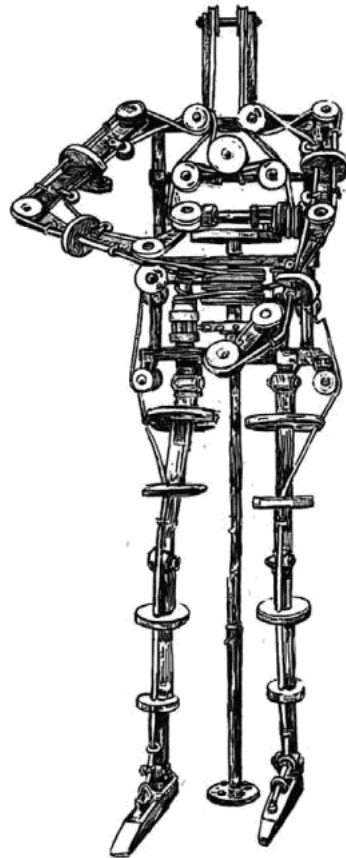


Британский словарь английского языка **Collins English Dictionary** назвал главным словом 2023 года AI (ИИ, искусственный интеллект)

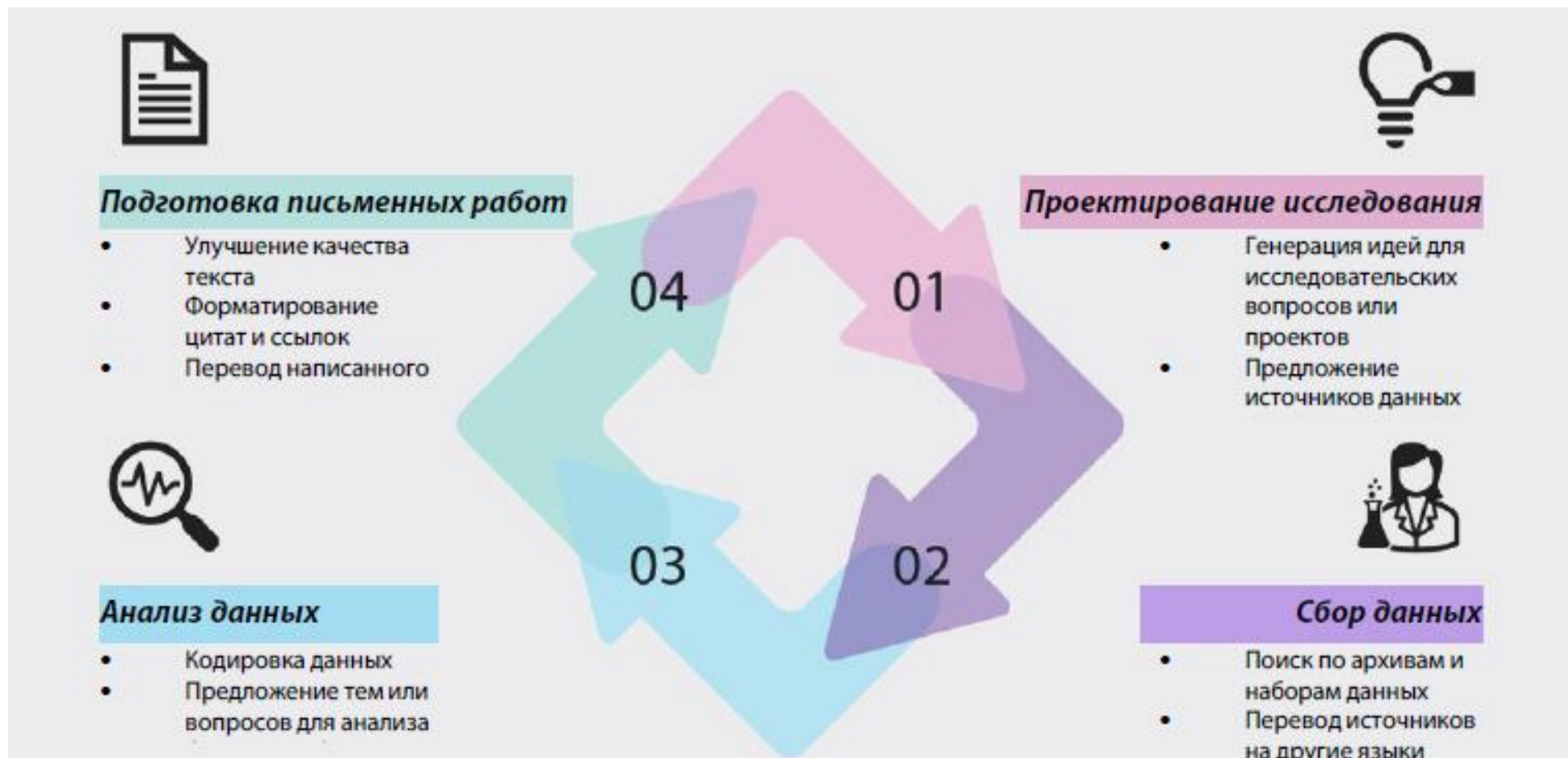
Определение понятия искусственного интеллекта

Искусственный интеллект относится к системам, которые демонстрируют интеллектуальное поведение, анализируя окружающую среду и предпринимая действия – с некоторой степенью автономности – для достижения конкретных целей*





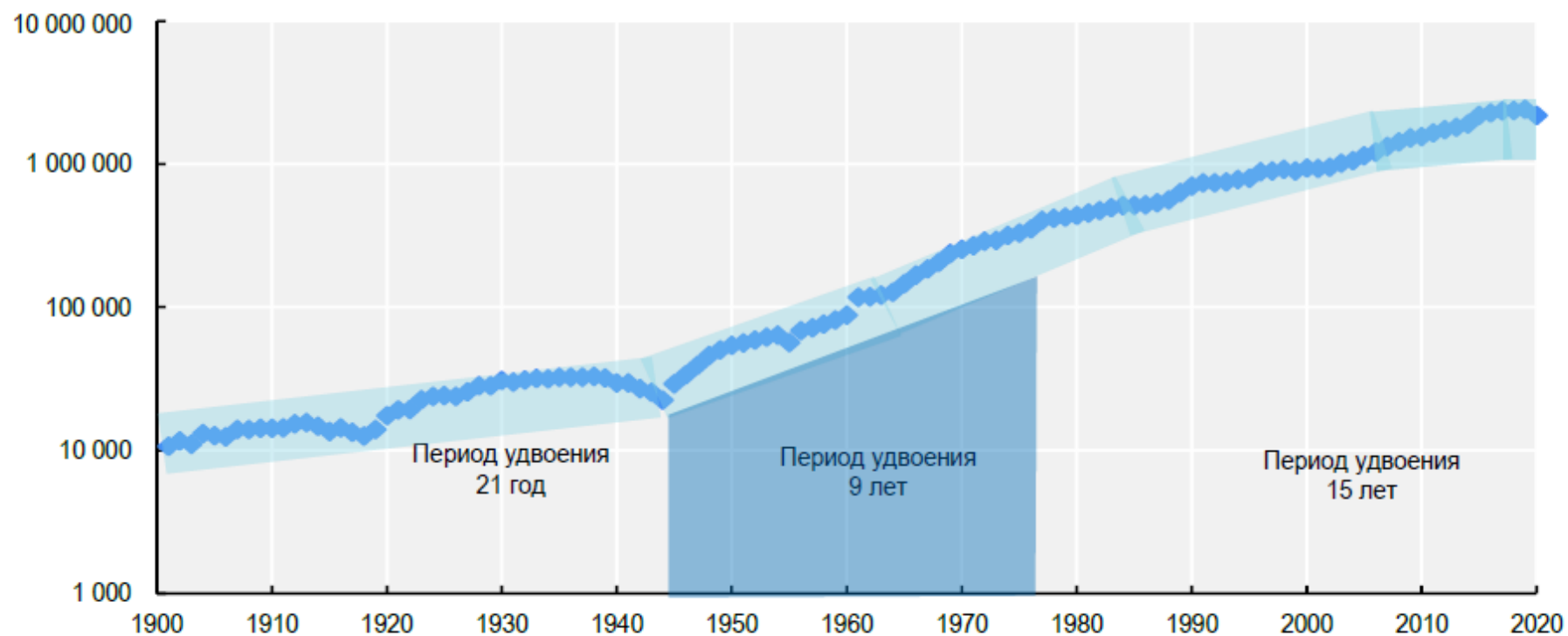
Британский математик Алан Тьюринг, 1950 год. Тест Тьюринга — метод определения наличия у машины интеллекта



Количество публикуемых научных работ

Все научные публикации

Публикации, логарифмическая шкала



7325977

публикаций *

Легкость, с которой исследователи могут загрузить десятки тысяч статей на свои компьютеры и распечатать сотни из них это одновременно и хорошо, и плохо.

1. В юридических рассуждениях адвокатам платят за просмотр миллионов документов в попытке найти доказательства по какому-либо делу.
2. В доказательной медицине исследователи просматривают медицинские публикации, чтобы собрать доказательства в поддержку определенной медицинской практики или явления.
3. В области разработки программного обеспечения систематические обзоры литературы стали стандартной процедурой в исследованиях.

Как правило, обзор литературы представляет собой в основном ручной процесс.

Для опытного ученого просмотр 5000 публикаций требует 40 часов непрерывной работы для простых тем и значительно больше времени для сложных тем *

Scopus – доступен поиск по авторам и журналам (196 млн зап., 1,9 млрд цит.)

WoS ([Master Journal List](#)) – просмотр журналов, доступна краткая информация об интересующем журнале (84 млн зап., 1,8 млрд цит.)

[Dimensions.ai](#) – доступен поиск по авторам и журналам (Crossref, PubMed, ORCID; 305 млн зап., 1,7 млрд цит.)

[Lens](#) – полноценный поиск по пристатейным спискам литературы (Microsoft Academic, Crossref, PubMed, ORCID, OpenAlex; 401 млн зап., 1,9 млрд цит.)

[Semantic Scholar](#) – поиск публикаций в научных журналах (Crossref, издатели, PubMed, репозитории, веб-сканирование; 211 млн зап., 2,4 млрд цит.) *

**снижает нагрузку на
поиск информации***

До 60%

**обеспечивает экономию
времени более***

80 часов

1. Повышение производительности за счет экономии времени.
2. Создание качественно новой работы, автоматической генерации гипотез.
3. Обнаружения дублирования.
4. Продвижение исследований доступных для неспециалистов.

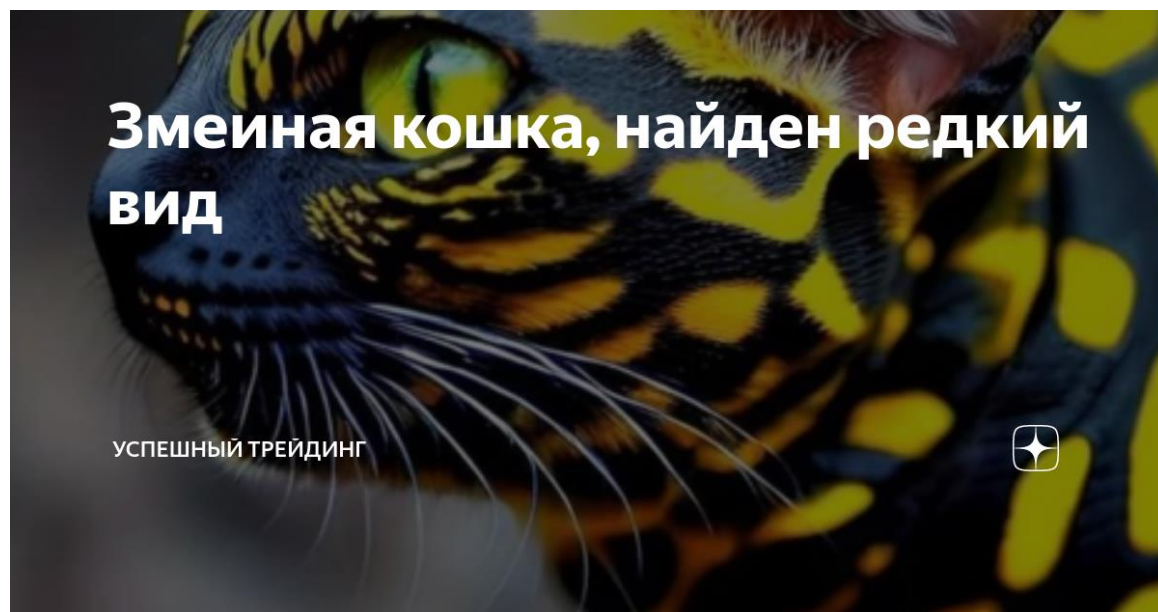
Возможности применения ИИ — от помощи в формулировании запросов ученых до возможности обрабатывать синонимы и сложные запросы.

Галлюцинировать – производить ложную информацию. Термин «галлюцинировать» стал одним из многих слов, которые в 2023 году обрели дополнительные значения в свете развития ИИ *

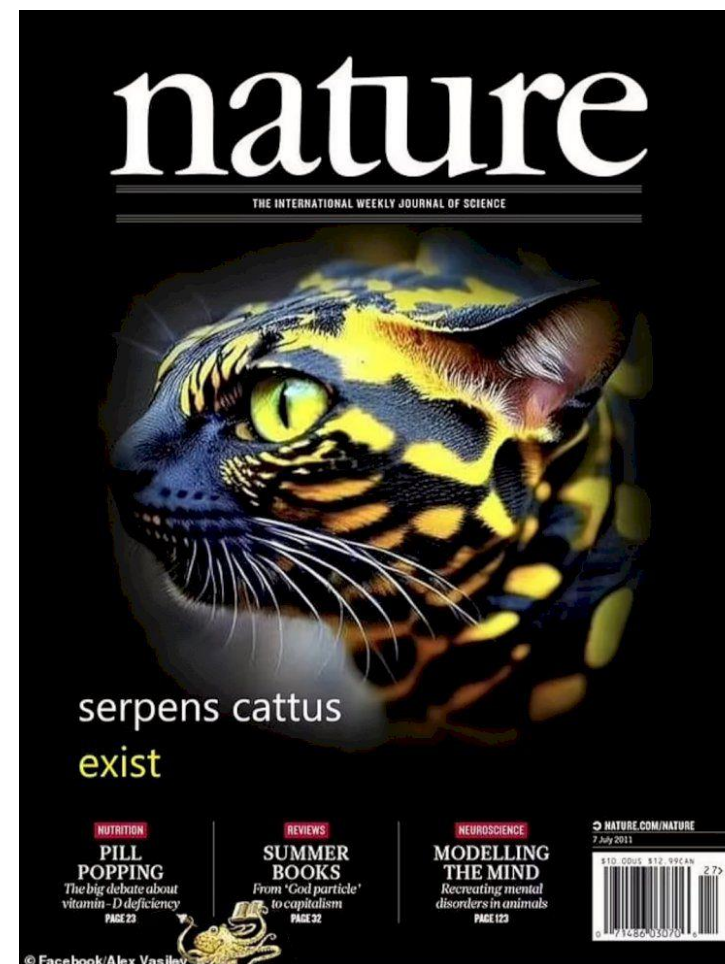
Кембриджский словарь английского языка **The Cambridge Dictionary** назвал главным словом 2023 года Hallucinate (галлюцинировать)



The screenshot shows the Cambridge Dictionary app interface. At the top, there is a navigation bar with a hamburger menu, the Cambridge Dictionary logo, a user profile icon, a globe icon, a dropdown arrow, and a search icon. Below the navigation bar, there is a banner featuring an illustration of a white airplane with orange wings flying against a dark blue background. The text on the banner reads: "The Cambridge Dictionary Word of the Year 2023 is...". Below the banner, a white card displays the word "hallucinate" in a large, bold, dark blue font. Underneath the word, it says "verb". A horizontal line separates the word from the definition: "When an artificial intelligence hallucinates, it produces false information." At the bottom right of the card, there is a small illustration of a white robot holding a globe.



Serpens cattus (змеиная кошка)





начало исследования (starting)

GigaChat, YandexGPT2, Garmini, Copilot, ChatGPT, Elicit, Concensus, Perplexity

анализ ссылок (chaining)

Inciteful, litmaps, Keenious, Research Rabbit

просмотр (browsing)

SciSpace, Elicit, Semantic Scholar, Lens

дифференцирование (differentiating)

Arrowsmith, Elicit

мониторинг (monitoring)

Semantic Scholar, Lens

извлечение (extracting)

SciSpace, Elicit, Myreader, Perplexity, ChatGPT

Начало исследования

GigaChat, YandexGPT2, Gemini, Copilot, ChatGPT, Elicit, Concensus, Perplexity

1. Возобновляемые источники энергии
2. Солнечная энергия
3. Ветровая энергия
4. Гидроэнергетика
5. Биоэнергетика

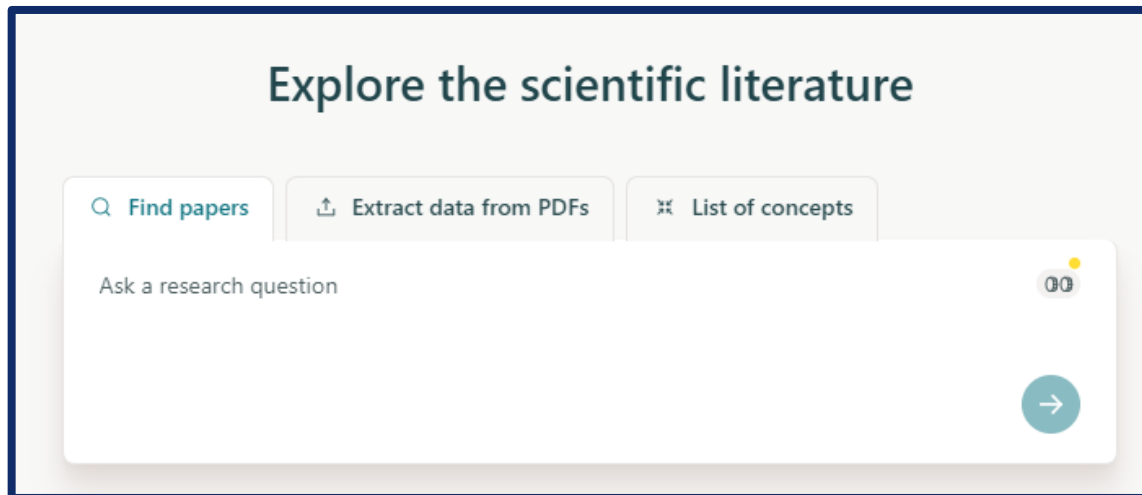
Пожалуйста, уточните, какого рода информацию вы хотели бы найти? Примеры могут включать статистические данные, исторические факты, научные исследования, технологии, использование в разных странах и т.д.

Prompt (промпт) – задача чат-боту

Заполненный запрос может выглядеть так:

«Мне нужно найти информацию о возобновляемых источниках энергии для исследовательской работы, которая будет включать пять ключевых слов».

Начало исследования



Elicit используется в основном для обзора литературы

Начало исследования

Rank	Prob	B-term
1	0.99	toll receptor
2	0.99	lipid nanoparticle
3	0.99	nipah virus
4	0.99	acinetobacter baumannii
5	0.99	aptamer
6	0.99	gold nanoparticle
7	0.99	pcsk9
8	0.99	conjugate vaccine
9	0.99	updated systematic review
10	0.99	emerging infectious disease
11	0.99	scoping review
12	0.99	pneumococcal conjugate
13	0.99	[-]pneumococcal conjugate vaccine
14	0.99	a scoping review
15	0.99	molecular docking
16	0.99	hbv
17	0.99	hpv
18	0.99	antimicrobial stewardship
19	0.99	rna therapeutic

AB literature	B-term	BC literature
covid	scoping review	vaccination effectiveness
<p>1: Livestreaming Technology and Online Child Sexual Exploitation and Abuse: A Scoping Review. 2024 Add to clipboard</p> <p>2: Exploring the COVID-19 Knowledge, Attitudes, and Practices (KAPs) in the Black Community: a Scoping Review. 2024 Add to clipboard</p> <p>3: Online interviews for qualitative health research in Africa: a scoping review. 2024 Add to clipboard</p> <p>4: A Scoping Review of Technology-Facilitated Gender-Based Violence in Low- and Middle-Income Countries Across Asia. 2024 Add to clipboard</p> <p>5: COVID-19 infection risk assessment and management at the Tokyo 2020 Olympic and Paralympic Games: A scoping review. 2024 Add to clipboard</p> <p>6: Hospices and Emergency Preparedness Planning: A Scoping Review of the Literature. 2024</p>	<p>1: Burden of respiratory syncytial virus disease across the lifespan in Australia and New Zealand: a scoping review. 2024 Add to clipboard</p> <p>2: The socioeconomic impact of equine epizootic lymphangitis in working equids in low and middle-income countries: A scoping review. 2024 Add to clipboard</p> <p>3: A Scoping Review on Malaria Prevention and Control Intervention in Fragile and Conflict-Affected States (FCAS): A Need for Renewed Focus to Enhance International Cooperation. 2024 Add to clipboard</p> <p>4: Malaria vaccine efficacy, safety, and community perception in Africa: a scoping review of recent empirical studies. 2024 Add to clipboard</p> <p>5: Increasing pneumococcal vaccine uptake in older adults: a scoping review of interventions in high-income countries. 2023 Add to clipboard</p>	

Start	A-Literature	C-Literature	B-list	Filter	Literature
Clipboard					
1.	37105230:	Harnessing and bioprospecting botanical-based herbal medicines against potential drug targets for COVID-19: a review coupled molecular docking studies.			found when browsing the A-literature using the B-term molecular docking
2.	37933071:	Molecular mechanism of the effect of Gegen Qinlian decoction on COVID-19 comorbid with diabetes mellitus based on network pharmacology and molecular docking : A review.			found when browsing the A-literature using the B-term molecular docking
3.	36683992:	The moderating effect of the COVID-19 pandemic on the mental wellbeing of health care workers on sustainable employability: A scoping review .			found when browsing the A-literature using the B-term scoping review
4.	36910342:	Knowledge, risk-perception, and uptake of COVID-19 prevention measures in sub-Saharan Africa: a scoping review .			found when browsing the A-literature using the B-term scoping review

Начало исследования

Rank	Job ID	C-query	nC	nAC	nTot	nR	pR
1	148587	Cells, Cultured [mh]	9999	3695	7675	1773	0.231
2	1485813	Epithelial Cells [mh]	9999	3842	8955	2024	0.226
3	148585	Blood Cells [mh]	9999	5481	10564	2386	0.226
4	1485821	Myeloid Cells [mh]	9999	1848	9251	2014	0.218
5	1485818	Hepatic Stellate Cells [mh]	4703	3	3799	824	0.217
6	1485829	Phagocytes [mh]	9999	1665	9145	1972	0.216
7	148582	Antibody-Producing Cells [mh]	9999	674	9009	1891	0.210
8	148583	Antigen-Presenting Cells [mh]	9999	221	8394	1725	0.206
9	148586	Bone Marrow Cells [mh]	9999	543	9398	1894	0.202
10	1485820	Muscle Cells [mh]	9999	128	6822	1345	0.197
11	1485823	Neoplastic Cells, Circulating [mh]	9999	4	7102	1350	0.190

A-Literature: **Covid 19**

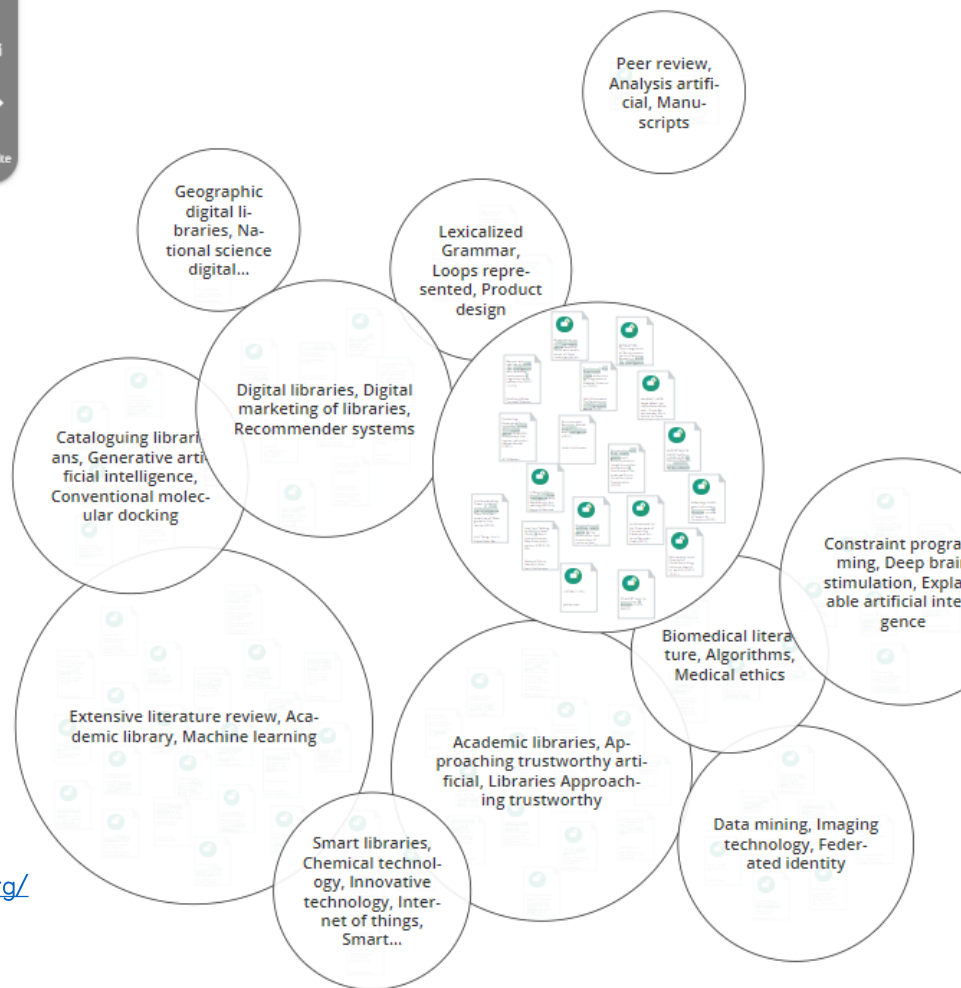
C-Literature: The list of individual MeSH terms within the MeSH Category of Cells

Knowledge Map of artificial intelligence in libraries

100 most relevant documents | Data source: BASE | Until 4 Apr 2024 | Document types | All lang | Data quality | [More information](#)

Начало исследования



The Use of Artificial Intelligence in Digital Marketing of the Library Industry in Ukraine: Foreign Experience and Prospects (2023-07-10)

Maranchak Nadiia
Digital Platform: Information Technologies in Sociocultural Sphere 6(1) 172-184
[link]: <https://zenodo.org/record/8131567>

The purpose of the article is to analyze the foreign experience of using artificial intelligence in the digital marketing of libraries, to outline the main directions, prospects and problems of applying these technologies in Ukraine, and to formulate recommendations for the successful implemen...

Area: [Digital libraries](#), [Digital marketing of libraries](#), [Recommender systems](#)

The Use of Artificial Intelligence in Digital Marketing of the Library Industry in Ukraine: Foreign Experience and Prospects ; Використання штучного... (2023-07-10)

Надія Маранчак
Цифрова платформа: інформаційні технології в соціокультурній сфері; Том 6 № 1 (2023); 172-184; Digital...
[link]: <http://infotech-soccult.knukim.edu.ua/article/view/283986>

The purpose of the article is to analyze the foreign experience of using artificial intelligence in the digital marketing of libraries, to outline the main directions, prospects and problems of applying these technologies in Ukraine, and to formulate recommendations for the successful implemen...

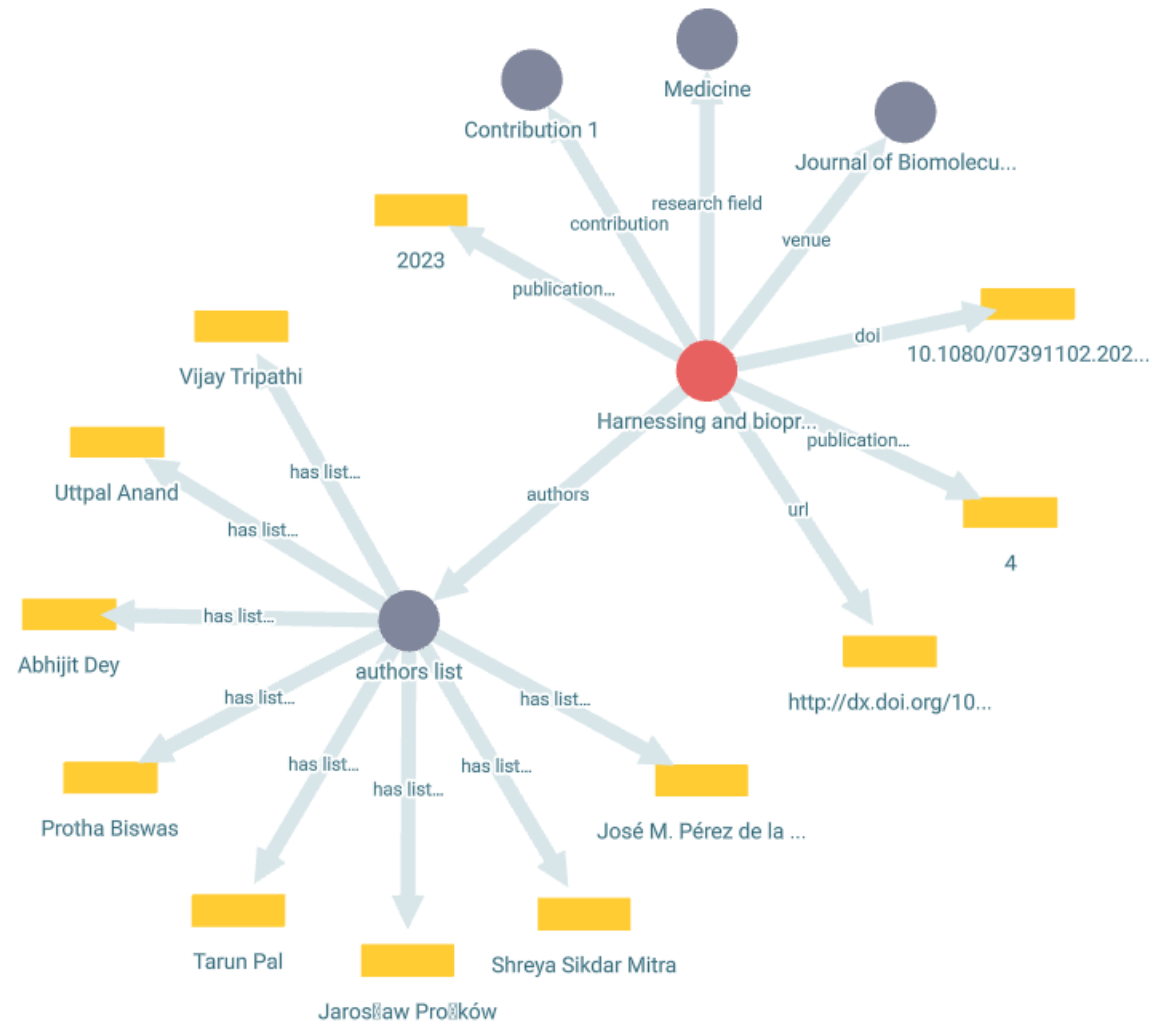
  

Area: [Digital libraries](#), [Digital marketing of libraries](#), [Recommender systems](#)

Анализ ссылок

ORKG семантически связывает различные научные идеи, подход методы и представляет их в машиночитаемом формате. Это позволяет облегчить доступ к последним достижениям.

DOI: [10.1080/07391102.2023.218763](https://doi.org/10.1080/07391102.2023.218763)



Анализ ссылок

Harnessing and bioprospecting botanical-based herbal medicines against potential drug targets for COVID-19: a review coupled molecular docking studies

DOI:
[10.1080/07391102.2023.2187634](https://doi.org/10.1080/07391102.2023.2187634)

<https://inciteful.xyz/>

Harnessing and bioprospecting botanical-based herbal medicines against potential drug targets for COVID-19: a review coupled molecular docking studies

Tarun Pal | Uttpal Anand | Shreya Sikdar Mitra | Protha Biswas | Vijay Tripathi | Jaroslaw Proćków
 | Abhijit Dey | Jose Manuel Pérez de la Lastra

Journal of Biomolecular Structure & Dynamics | [10.1080/07391102.2023.2187634](https://doi.org/10.1080/07391102.2023.2187634)

1	105	2023	Yes	26,049	291,062	2
Cited By	Citing	Published	Open Access	Papers in Graph	Citations in Graph	Graph Depth

Links: [Publisher](#) | [Full Text from LibKey](#)

Каждая сеть и данные в ней уникальны для вашего поиска и основаны на документах, которые вы ей предоставили, поэтому вы можете быть уверены, что получите максимально релевантные результаты.

Add columns (3) ▾
Type ▾
Sort by: ▾
Export ▾

Files (16)	Summarized Abstract	Methods Used	Results
<p>📄 Scholar Home / My Library / All librarian / s13643-020-01324-7.pdf</p> <p>Last open</p>	<div style="border: 1px solid #ccc; padding: 5px;"> <p style="text-align: center;">Systematic Reviews</p> <p style="text-align: center;">RESEARCH Open Access</p> <h3 style="text-align: center;">Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer</h3> <p style="text-align: center;">Amy Y. Isou¹, Jonathan R. Treadwell, Eileen Erinoff and Karen Schoelles</p> <div style="border: 1px solid #000; padding: 5px; margin-top: 10px;"> <p>Abstract</p> <p>Background: Improving the speed of systematic review (SR) development is key to supporting evidence-based medicine. Machine learning tools which semi-automate citation screening might improve efficiency. Few studies have assessed use of screening prioritization functionality or compared two tools head to head. In this project, we compared performance of two machine learning tools for potential use in citation screening.</p> <p>Methods: Using 9 evidence reports previously completed by the ECR Institute Evidence-based Practice Center team, we compared performance of Abstrackr and EPPI Reviewer, two off the shelf citations screening tools, for identifying relevant citations. Screening prioritization functionality was tested for 3 large reports and 6 small reports on a range of clinical topics. Large report topics were imaging for pancreatic cancer, indoor allergen reduction, and inguinal hernia repair. We trained Abstrackr and EPPI Reviewer and screened all citations in 10% increments. In Task 1, we inputted whether an abstract was ordered for full text screening; in Task 2, we inputted whether an abstract was included in the final report. For both tasks, screening continued until all studies ordered and included for the actual reports were identified. We assessed potential reductions in hypothetical screening burden (proportion of citations screened to identify all included studies) offered by each tool for all 9 reports.</p> <p>Results: For the 3 large reports, both EPPI Reviewer and Abstrackr performed well with potential reductions in screening burden of 4 to 49% (Abstrackr) and 9 to 60% (EPPI Reviewer). Both tools had markedly poorer performance for 1 large report (inguinal hernia), possibly due to its heterogeneous key questions. Based on McNemar's test for paired proportions in the 3 large reports, EPPI Reviewer outperformed Abstrackr for identifying articles ordered for full text review, but Abstrackr performed better in 2 of 3 reports for identifying articles included in the final report. For small reports, both tools provided benefits but EPPI Reviewer generally outperformed Abstrackr in both tasks, although these results were often not statistically significant.</p> <p>Conclusions: Abstrackr and EPPI Reviewer performed well, but prioritization accuracy varied greatly across reports. Our work suggests screening prioritization functionality is a promising modality offering efficiency gains without giving up human involvement in the screening process.</p> <p>Keywords: Machine learning, Citation screening, Text mining, Abstrackr, EPPI Reviewer, Screening prioritization, Methodology, Screening burden, Efficiency</p> </div> </div>	<div style="border: 1px solid #ccc; padding: 5px;"> <p style="text-align: center;">Copilot</p> <p style="text-align: center;">Standard High Quality ⚙️</p> <p style="text-align: center;">Как алгоритмы машинного обучения отдадут приоритет цитатам?</p> <p style="text-align: right;">Yesterday</p> <div style="border: 1px solid #000; padding: 5px; margin-top: 10px;"> <p>“Алгоритмы машинного обучения приоритизируют цитаты следующим образом:”</p> <ul style="list-style-type: none"> Инструменты машинного обучения используют функцию приоритизации скрининга для выявления релевантных ссылок и повышения эффективности систематических обзоров. [1] Эти инструменты генерируют алгоритмы прогнозирования и отдадут приоритет непроверенным ссылкам, отсортировав их от наиболее релевантных к наименее релевантным. [2] Точность приоритизации алгоритмов машинного обучения различается в разных отчетах, причем некоторые отчеты показывают более высокую производительность, чем другие. [1] Алгоритмы машинного обучения позволяют снизить нагрузку на проверку за счет раннего выявления релевантных ссылок, что может привести к уменьшению количества ссылок, подлежащих проверке. [3] Различные инструменты машинного обучения могут отличаться эффективностью при идентификации статей, заказанных для полнотекстового рецензирования, и статей, включенных в итоговый отчет. [3] Алгоритмы машинного обучения могут повысить эффективность процесса отбора без исключения части цитата. [1] </div> <p>Explain Abstract of this paper in 2 lines. What are the contributions of this paper, Find I</p> <p>Write your question...</p> <p>Σ MATH</p> </div>	<p>ЭКОНОМИТ</p> <p>Поиск по т</p>
<p>📄 s13643- Last open</p>			
<p>📄 s12874- Last open</p>			

Языковые модели могут преобразовать процесс исследования тремя способами:

1. повысить производительность за счет экономии времени,
2. создать качественно новую работу и
3. сделать исследования доступными для неспециалистов.

Спасибо за внимание



@FEDOROV DAYS