элпаб.
ОБРАЗОВАНИЕ

Сайт журнала в эпоху ИИ: индексация, оптимизация, целостность смысла



30 октября 2026 г.,

Образовательная программа Элпаб

Зельдина Марина Михайловна,
НЭИКОН, Элпаб

zeldina@neicon.ru

NEICON
ЭЛЕКТРОННАЯ ИНФОРМАЦИЯ

ПЛАТФОРМА ЭЛПАБ

Что такое Элпаб

- Элпаб был создан при участии Министерства науки и высшего образования по Госконтракту 14.514.11.4028 в 2013 г., получил признание и продолжает развиваться, находя специальные решения многочисленных ситуаций совмещения международных требований к научным публикациям с российскими реалиями.
- Элпаб — это платформа комплексной поддержки и сопровождения научного журнала, которая обеспечивает соответствие всем требованиям, пожеланиям и условиям международных баз любого уровня, ВАК, Белого списка, обладая всеми технологическими особенностями, присущими современному высокоуровневому журналу.
- Всего с помощью Элпаб проведено более 13 000 индексаций различных журналов с различными реферативными и полнотекстовыми, российскими и международными базами данных и индексами цитирования.

Scopus	124
WoS	21
DOAJ	168
CNKI	350
РИНЦ	≈700
RSCI	79

Составляющие Элпаб



Электронная версия научного журнала и осуществление всех технологических процессов



Методическая поддержка редакции и сопровождение журнала



Производство научного журнала

Элпаб в цифрах

- 76 специалистов — методисты, эксперты, библиографы, программисты, юристы, администраторы, дизайнеры
- 900 журналов с 2013 г.
- 23 портала организаций
- 15 репозиториев
- Не менее 300 выпусков в месяц
- Более 35 000 опубликованных статей в год

Цели и задачи Элпаб



- Привести электронную версию журнала в соответствие с современными общепринятыми нормами и правилами путем структуризации контента и введения обязательных разделов
- Вывести журнал на международный уровень путем индексации в международных индексных и реферативных БД
- Освободить редакцию журнала от всех технических вопросов и задач, связанных с работой сайта и обработкой контента
- Дать любые рекомендации и ответы на вопросы, связанные с изданием научной периодики
- Помочь подготовить, оформить и опубликовать научный контент в полном соответствии с общепринятыми стандартами
- Оперативно реагировать на изменения издательской и публикационной среды
- Встроить журнал в систему отчетности о научных результатах (публикациях)
- Сориентировать редакцию в современном правовом поле

САЙТ ЖУРНАЛА В ЭПОХУ ИИ

С чего все началось



элпаб.
ОБРАЗОВАНИЕ

Индексация сайта журнала

Критически важно обеспечить индексацию сайта журнала.

Благодаря видимости контента журнала для поисковых роботов контент журнала:

- доступен в крупнейших поисковых системах (Яндекс, Google);
- доступен для индексации специализированными научными поисковиками;
- доступен для индексации роботами Scopus

Доля ботов < 5%

Индексация журнала, 2025-2026

ИИ-роботы

- Научные AI: Elicit, Scholarcy, Prism, CNKI, PaperPal
- Коммерческие LLM, т.е. Большие языковые модели (Large Language Model): OpenAI GPT, Claude, Gemini, DeepSeek и др.
- Корпоративные боты: Microsoft Research, Google DeepMind, NVIDIA (сбор данных для доменных моделей) и др.
- Национальные ИИ: «GigaChat» (РФ), YandexGPT (РФ), «Ernie» (Китай), «Tencent AI» (Китай), «Falcon» (ОАЭ) и др.
- Агрегаторы: Perplexity AI, You.com, Bing Copilot и др.
- Маргинальные: Боты мелких организаций, студенческие, временщики

Слайд из презентации М. Митрофанова

Хороший робот, плохой робот

Хорошие роботы

- Обеспечивают присутствие контента в выдаче крупнейших ИИ-платформ

Плохие роботы

- Не приносят пользы журналу
- Создают нагрузку на сайт журнала, иногда критическую (и сайт падает)

Готовых алгоритмов, которые помогали бы отличать хороших роботов от плохих – нет. Иногда плохие роботы притворяются хорошими. Важно правильно их распознать.

Каждый администратор сайта оценивает угрозу и создает ловушки на основе своего практического опыта отражения атак.

Все ИИ-роботы портят статистику посещаемости сайта

Раньше по данным Яндекс.Аналитики и других сервисов можно было в моменте получить информацию о читателях журнала:

- Из каких стран приходят люди;
- Какие статьи им интересны;
- На каких страницах сайта они задерживаются чаще всего.

На основании этого анализа можно было управлять контентом журнала и выбирать наиболее интересные аудитории темы.

С появлением ИИ стало гораздо сложнее получать информацию об аудитории журнала.

Я – робот

«Нечеловеческие» характеристики просмотров:

- Просмотрено много страниц за ограниченное время;
- Глубина просмотров очень низкая;
- Таких просмотров очень много, в общей статистике они видны как всплески на графике

В этот момент важно проанализировать, какие именно боты заходили на сайт журнала, заблокировать «плохих» и предпринять меры, чтобы ограничить дальнейшие посещения, не приносящие пользы журналу.

Иногда даже посетители-люди при более внимательном рассмотрении оказываются посетителями-роботами

Что видно:

- **11.10** — пик ботов (983 визита, отказы 96%, время на сайте 0,8 сек — чистые краулеры/сканеры). Но аномалия не только в них: даже сегмент «люди» в тот день — 2 255 визитов с отказами 95% и временем 7,7 сек. Это больше похоже на нераспознанных ботов, чем на реальную аудиторию.
- **Фоновый бот-шум** — 60–160 визитов в день в большинство дней; 13–15 октября Метрика роботов не зафиксировала вовсе, 16–17 октября снова всплеск.
- Аномально длинные сессии «людей» 10.10 (171 сек) — возможно, зависшие вкладки, но это уже другое.

Характеристики флуда (прямые заходы):

- **Время:** 1 796 визитов за один час 22:00–22:59, ещё 172 в 23:00 и 51 в 21:00 — то есть ~2 200 хитов за два часа, остальной день чистый
- **Точка входа:** 2 099 из 2 184 — главная /, один просмотр, 98,7% отказов
- **Фингерпринт:** ровный разброс по «свежим» комбинациям ОС/браузера — iOS 18/Safari (592), Android 16/Chrome Mobile (537), macOS Sonoma/Chrome (504), Win10/Chrome (405) — у всех 100% отказов и 0 секунд на сайте. Настоящая аудитория так равномерно по железу не распределяется

Диагноз: генератор трафика бил в главную страницу в ночь воскресенья 11.10 (21:00–23:59 по Москве), подделывая User-Agent'ы актуальных браузеров. Метрика поймала только 983 робота, ещё ~2 100 прошли как «люди». К сайту это вреда не несёт, но портит статистику.

Результаты поисковой выдачи

Раньше

спектроскопия комбинационного рассеяния углер…

Поиск | Алиса AI | Картинки | Видео | Карты | Товары | Финансы | Квартиры

Московский Государственный Университет
kge.msu.ru › education/prak_17.pdf
Определение диаметра одностенных **углеродных нанотрубок** методом **спектроскопии комбинационного рассеяния**. Харланов А.Н.

PDF Посмотреть

Исследование спектров **комбинационного**...
fundamental-research.ru › article/view...
Спектроскопия комбинационного рассеяния (КР) является одним из инструментов исследования структуры графеновых слоев **углеродных нанотрубок** (**УНТ**). В данной работе изучены спектры КР синтезированных методом...

Анализ атомной структуры **углеродных нанотрубок**...
cyberleninka.ru › article/n/analiz-atomnoy-...
В данной работе представлен экспериментальный метод определения структуры **углеродных нанотрубок**, основанный на комбинировании **спектроскопии комбинационного рассеяния** света, электронной микроскопии...

У глеродные материалы, в силу своих многочисленных...
j-analytics.ru › files/article_pdf/3/article_3064_...
Спектроскопия комбинационного рассеяния – перспективный метод исследования **углеродных** наноматериалов. **Углеродные нанотрубки** – это продукт

Сейчас

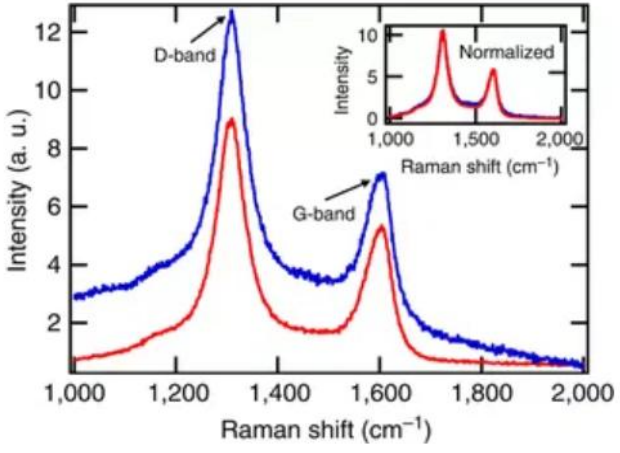
спектроскопия комбинационного рассеяния углер…

Поиск | Алиса AI | Картинки | Видео | Карты | Товары | Финансы | Квартир

Быстрый ответ Алисы AI

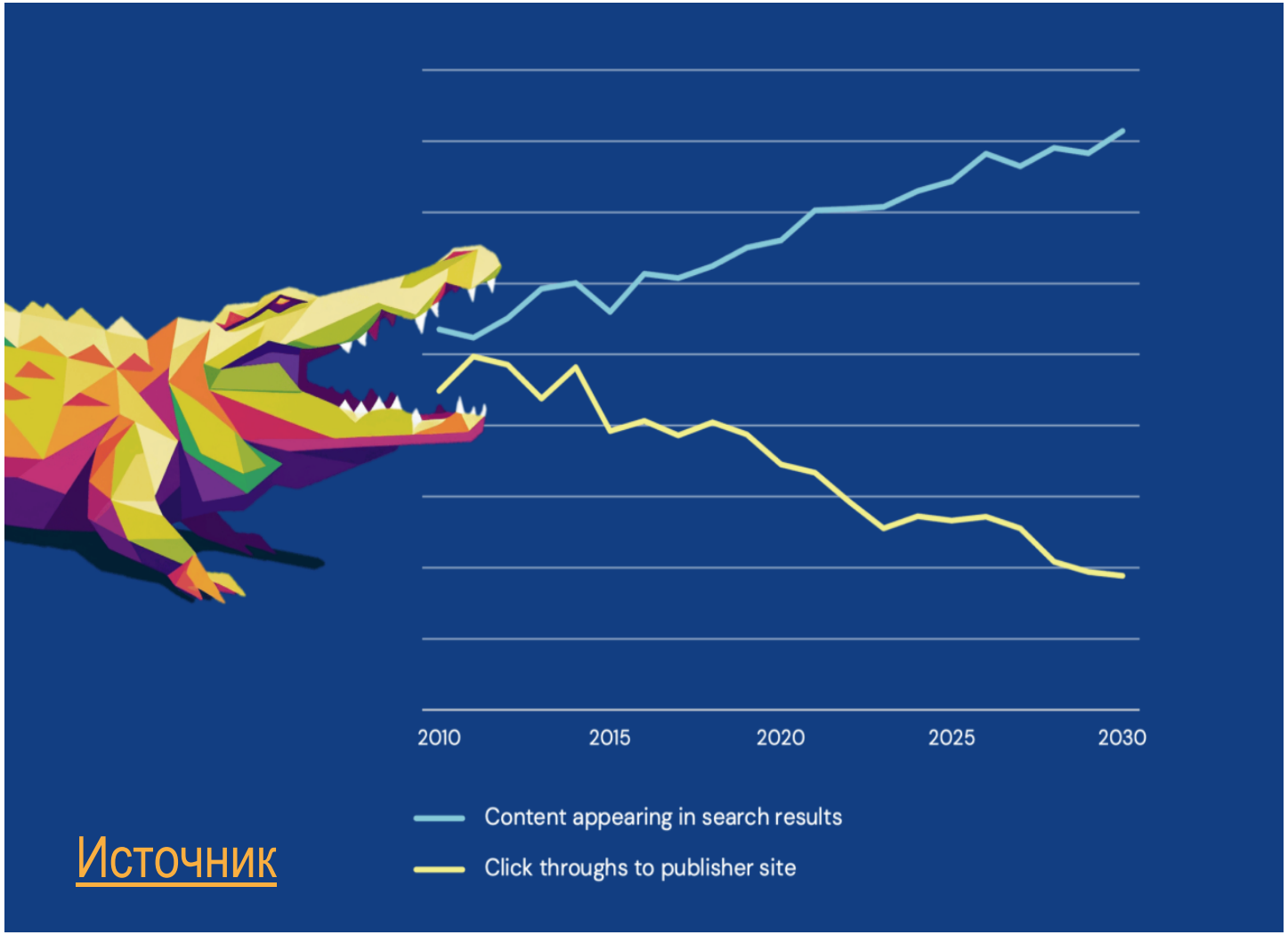
Спектроскопия комбинационного рассеяния (КР, или рамановская спектроскопия) — один из ключевых методов для изучения углеродных нанотрубок (УНТ). Она позволяет получать информацию об атомной структуре, колебательных модах и наличии дефектов.

nuclphys.sinp.msu.ru | cyberleninka.ru



Принцип метода

«Эффект крокодила»: поисковых запросов больше, переходов на сайт меньше



Источник

«Эффект крокодила» - это разрыв между количеством показов вашего контента в поисковых системах и числом переходов к полному тексту материала на вашем сайте

Источник

Потеря первоисточника

При традиционном поиске владельцами информации были научные журналы - статья-первоисточник была видна.

ИИ добавляет собственный фильтр. Система определяет:

- какие работы показывать в ответ на запрос
- какие результаты выделять;
- какие точки зрения опускать

Этот ИИ-фильтр не прозрачен. Мы не знаем, по каким критериям отбирается и обрабатывается информация, которую роботы получают на сайтах журналов.

Потеря первоисточника

- ИИ становится посредником между источником информации и пользователем. Поисковые роботы не только считывают информацию, эта информация затем обрабатывается и интерпретируется ИИ-агентами.
- В ответ на свой поисковый запрос человек получает ответ, сгенерированный ИИ-агентом. В разное время разные ИИ-агенты будут давать разные ответы.
- Комбинация «ответ, сгенерированный ИИ + ссылка на источник» не гарантирует корректности информации (потому что ответ будет один, а источников к нему – 5, и в них могут содержаться противоречащие сведения).
- От понимания важности обращения к первоисточнику будет зависеть, перейдет ли читатель на сайт журнала, чтобы проверить информацию.

Zero-click search / Нулевой клик

Концепция «нулевого клика» (zero-click readership) описывает ситуацию, когда содержание исследования доводится до аудитории, интерпретируется и даже используется повторно без обращения читателя к первоисточнику.

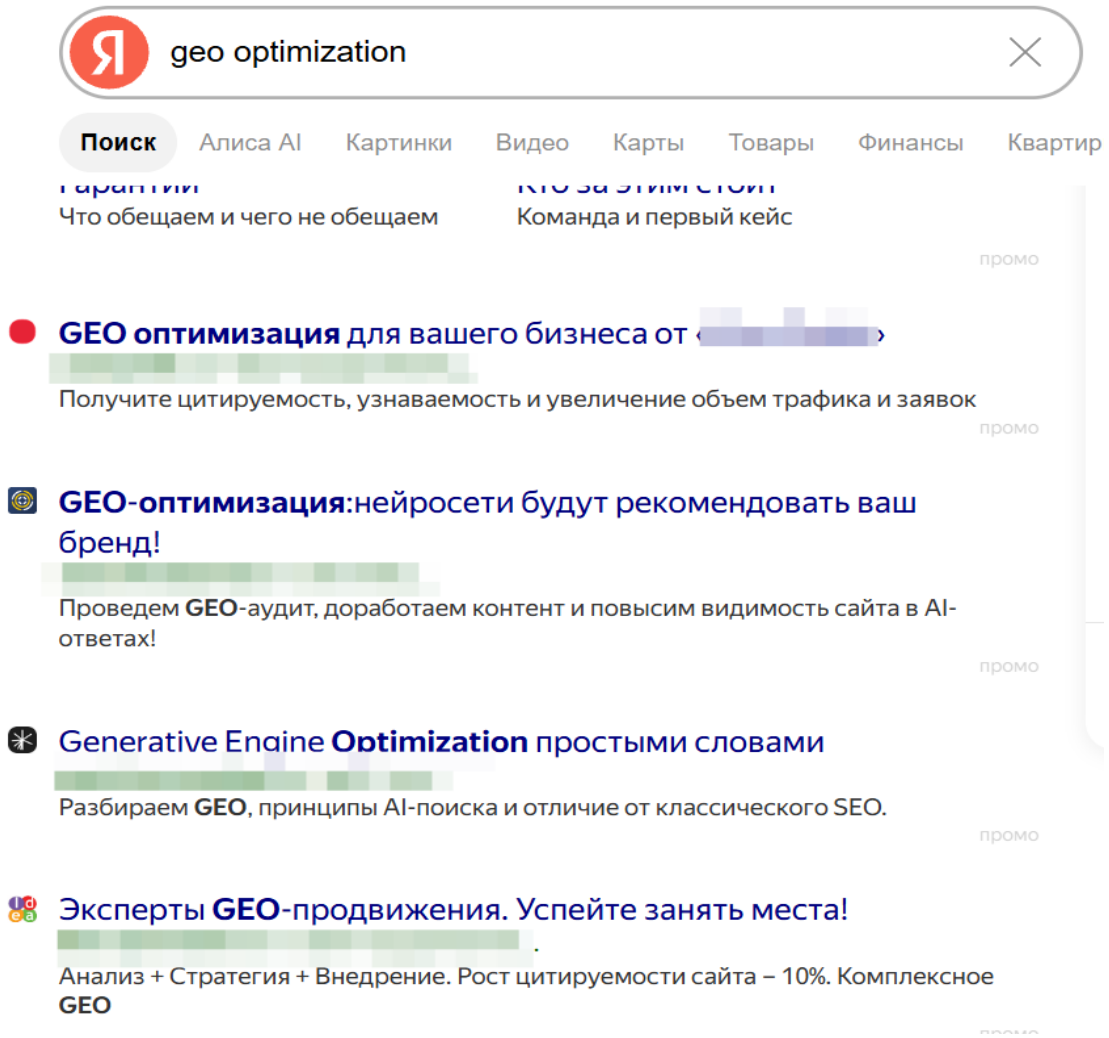
Раньше: Поиск -> Просмотр списка результатов -> клик на перспективные статьи -> чтение/просмотр этих статей -> и цитирование

Сейчас: Поиск -> Чтение сгенерированного ИИ обзора -> Переход к следующему вопросу

Источник

Можно ли вообще оптимизировать сайт под ИИ-поиск?

Реклама говорит, что да



Generative Engine Optimization (GEO) - это оптимизация контента под генеративные ИИ и нейропоиски

элпаб. ОБРАЗОВАНИЕ

Гомеопатия от IT

- Понятных и измеримых критериев «хорошей» GEO-оптимизации нет
- Известные исследования и рекомендации ограничиваются общими размытыми критериями, которым статья в научном журнале соответствует по умолчанию

Как оптимизировать то, что уже есть?

Улучшит качество индексации, но потребует больших усилий:

- Улучшение машиночитаемости PDF

На практике достичь 100% машиночитаемости в настоящий момент сложно:

- PDF не создавался для того, чтобы быть машиночитаемым. Это формат для тех, кто читает глазами, а не алгоритмами;
- верстальщики работают в разных программах и даже в разных версиях одной программы;
- не все верстальщики по умолчанию умеют обогащать PDF нужными метаданными;
- нет критериев для определения «хорошо обогащенного» PDF;
- но есть понимание, что машиночитаемость PDF может ухудшить.

Что ухудшает «машиночитаемость» PDF

- Верстка метаданных в две колонки
- Буквицы в начале статьи
- Многочисленные спецсимволы, включая неразрывные пробелы и неразрывные дефисы
- Таблицы со сложной версткой
- Многоуровневые колонтитулы и сноски
- Отсутствие подразделов в полном тексте статьи
- Отсутствие маркеров начала и конца полного текста статьи

Результат: LLM выдёргивает строки в неправильном порядке, путает подписи к колонкам, добавляет лишние символы/строки

Как оптимизировать то, что уже есть?

Улучшит индексацию, потребует меньше усилий:

- Размещение полных текстов в HTML;
- Размещение дополнительных материалов, связанных со статьей, на странице статьи (хайлайты, графические аннотации, видео).

Любые полнотекстовые материалы - это большой плюс, потому что ИИ-краулеры собирают с сайтов все. Они делают более точным индексирование.
HTML - последовательный набор строк, создающий контент. Последовательная индексация этих строк позволяет корректно проиндексировать текст без потери смысла

При индексации PDF набор строк может меняться, из-за этого проиндексированный текст может потерять часть смысла.

Делай, что должно – и будет нормально

«В наиболее выигрышной ситуации оказываются организации, которые часто готовят экспертные текстовые материалы. Исследователи из Принстонского университета, Технологического института Джорджии, Индийского технического института в Дели создали бенчмарк GEO-bench на основе 10 тыс. поисковых запросов на разные темы. В экспериментах с gpt-3.5-turbo и Perplexity выяснилось, что добавление цитат, статистики и ссылок на авторитетные источники повышает видимость материала на 35–40%».

Источник

Plain Language Summary

Plain Language Summary (текст на простом языке) – дополнительная аннотация к тексту научной статьи, написанная языком, понятным широкой публике. Прямая связь с более широкой аудиторией журнала – без посредников в виде пресс-служб и журналистов. Некоторые крупные издательства имеют рекомендации по созданию Plain Language Summary.
Источник

Ключевая рекомендация - создание "готового к ответу" контента (Answer-Ready Content). Издателям необходимо создавать структурированный контент на простом, понятном языке, который ИИ сможет легко интерпретировать и цитировать. Это и есть та самая оптимизация под генеративные движки - Answer-Engine Optimization (AEO) или Generative Engine Optimization (GEO).

Источник

Plain Language Summary

Springer Nature

Plain Language Summary включены в платный «медиа-пакет» автора, генерируются только по запросу, обязательно должны быть проверены автором.

Wolters Kluwer

Ovid AI Article Summary – аннотация генерируется автоматически, имеет стандартную структуру, доступна для статей в открытом доступе с полным текстом в HTML

Источник

Потенциальный стандарт

Подобно тому, как электронное распространение контента перевернуло многие структуры и системы обмена печатным контентом, роль систем ИИ потребует переосмысления всего, от доступа, цитирования и отслеживания использования до финансовых моделей, поддерживающих весь рынок. В то время как процесс перехода от печатного к электронному формату занял три десятилетия, темпы нашей адаптации к новому миру ИИ должны значительно ускориться.

TRACE: Trusted Retrieval & Attribution for Content Ecosystems - это инициатива Springer Nature, COUNTER, Ithaka S+R и NISO направленная на объединение членов научно-издательского сообщества и сообщества ИИ на благо всех участников экосистемы научных коммуникаций.

Инициатива стартовала с финансированием трех параллельных проектов, направленных на решение ключевых проблем в научной коммуникации с использованием ИИ:

- как агенты ИИ находят и получают доступ к контенту;
- как сохраняется происхождение и авторство на протяжении всего жизненного цикла контента;
- как можно последовательно измерять и сообщать об использовании ИИ.

Источник

Итоги

- Качество журнала не определяет качество его индексации. Если добросовестные журналы не будут задумываться о том, что проблема существует, то недобросовестные будут пытаться ее решить прямо сейчас
- Отсутствие четких рекомендаций – основная проблема
- Более широкое распространение HTML и XML – наряду с PDF
- Обогащение сайтов дополнительным контентом

элпаб.
ОБРАЗОВАНИЕ

Адресные вебинары о работе с генИИ

Элпаб проводит адресные вебинары по работе с генИИ для редакций журналов и ученых

Для связи: zeldina@neicon.ru
www.elpub.ru









Образовательная программа Элпаб

Ежегодная аудитория Образовательной программы более 1000 уникальных слушателей

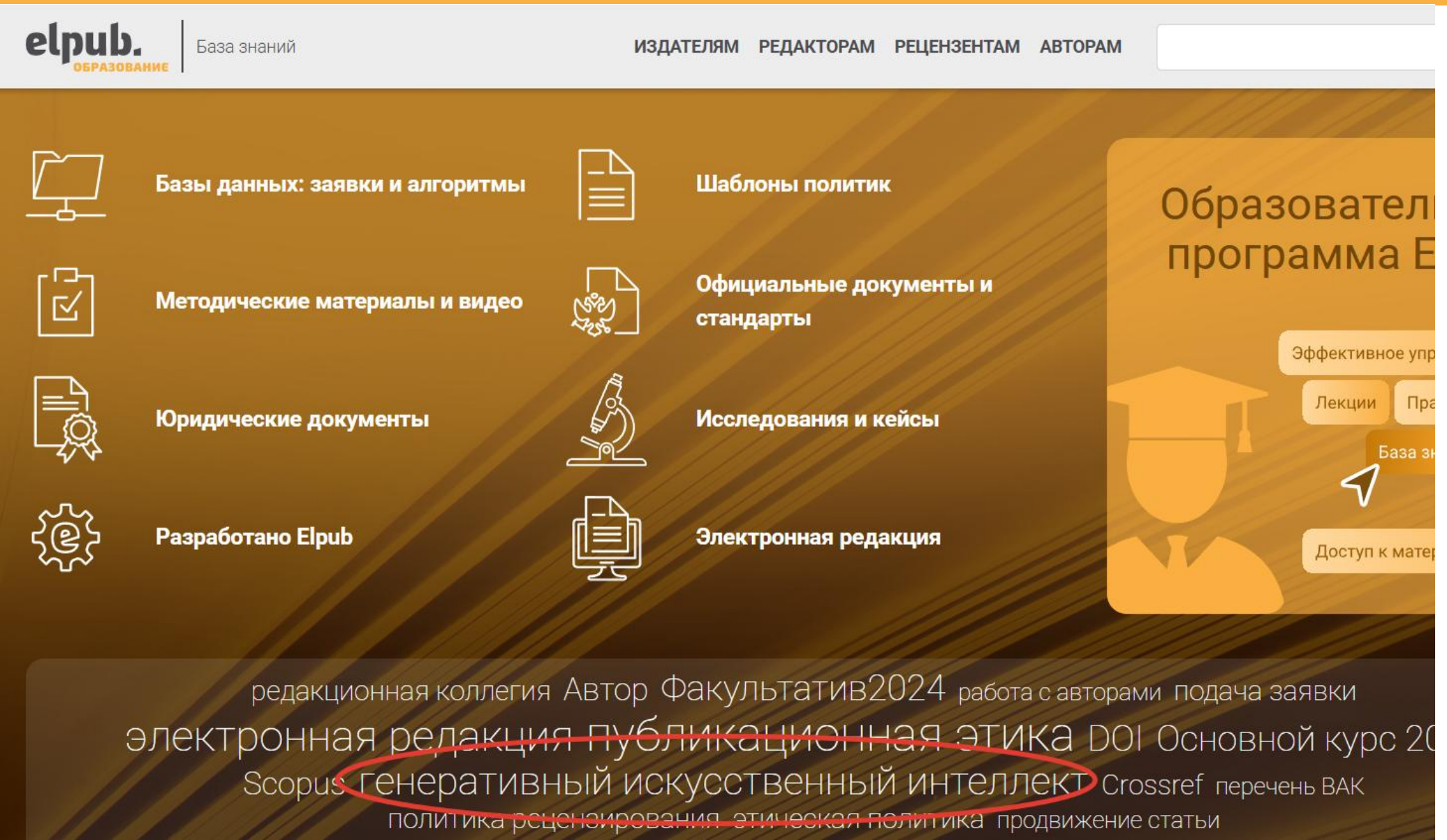
Ключевой образовательный результат

Мы ожидаем, что участники образовательной программы улучшат качество издания своего журнала и сделают его продвижение более эффективным, применяя принятые в издательском сообществе принципы, правила и стандарты.

После завершения программы участники смогут:

- Осуществлять эффективное управление научным журналом
- Разрабатывать и внедрять ключевые политики журнала
- Использовать все возможности Элпаб
- Управлять кадровыми вопросами
- Продвигать журнал
- Интегрировать инновации в издательский процесс
- Применять полученные знания на практике
- Оставаться в курсе последних тенденций

База знаний Элпаб: https://edu.elpub.ru/



Книга Elpub

Развитие и продвижение научного журнала: от теории к практике

- Разработка политики журнала
- Стратегии продвижения
- Взаимодействие с авторами
- Управление редакцией
- Технологии генИИ в работе научного журнала



+ листы самооценки

+ пошаговые рекомендации

[База знаний Elpub]

Все, что нужно знать о платформе Элпаб

По этому QR-коду можно скачать презентацию о платформе Элпаб

«Платформа Elpub. Комплексная поддержка и сопровождение научного журнала»

элпаб.
ОБРАЗОВАНИЕ

Спасибо за внимание !

www.elpub.ru

NEICON
ЭЛЕКТРОННАЯ ИНФОРМАЦИЯ

