

<https://doi.org/10.29188/2712-9217-2025-11-3-32-37>

# Использование AI-инструментов для повышения качества научных публикаций через автоматизированный анализ препринтов статей

Г.М. Жигулин<sup>1,2</sup>

<sup>1</sup> Кафедра цифровой медицины, Первый Московский государственный медицинский университет им. И.М. Сеченова (Сеченовский Университет), Москва, Россия

<sup>2</sup> ГБУЗ «Морозовская ДГКБ ДЗМ», Москва, Россия

**Контакт:** Жигулин Г.М., [glebzhigulin.pp@gmail.com](mailto:glebzhigulin.pp@gmail.com)

## Аннотация:

**Введение.** В условиях экспоненциального роста медицинских знаний и публикационной активности критическим фактором становится скорость доведения научных результатов до профессионального сообщества. Оценка длительности процесса рецензирования и публикации в биомедицинских журналах показывает, что значительные задержки часто вызваны не отсутствием научной новизны, а формальными несоответствиями и техническими ошибками в рукописях.

**Цель.** Разработка и обоснование внедрения автоматизированной системы на основе искусственного интеллекта (AI) для предварительного анализа препринтов (проект Ptolemaea), направленной на снижение процента формальных ошибок и ускорение публикационного цикла.

**Материалы и методы.** Исследование базируется на принципах Data-Centric AI, наукометрии и компьютерной лингвистике (NLP). Применяемая NLP-модель используется не для генерации текста, а для глубокого структурного и стилистического анализа: выявления нарушений логики IMRAD, несоответствия форматирования, проверки статистических показателей и соблюдения критериев конкретных журналов.

**Результаты.** Анализ публикационной активности показывает, что средний срок от подачи до принятия статьи варьируется от 50 до 276 дней. При этом до 93,2% ошибок, приводящих к доработкам или отказам, исходят от авторов (оформление, статистика, неполные данные). Представленная технология позволяет авторам выявлять данные недостатки на этапе препринта.

**Выводы.** Автоматизированный анализ препринтов является эффективным практическим инструментом, снижающим порог вхождения в научную деятельность для молодых ученых и устраняющим рутинизацию процесса подготовки рукописи для опытных врачей-исследователей. Внедрение подобных систем способствует повышению качества научной коммуникации в сфере здравоохранения.

**Ключевые слова:** искусственный интеллект; научные публикации; препринты; NLP; рецензирование; медицинская статистика; Ptolemaea; наукометрия.

**Для цитирования:** Жигулин Г.М. Использование AI-инструментов для повышения качества научных публикаций через автоматизированный анализ препринтов статей. Российский журнал телемедицины и электронного здравоохранения 2025;11(3):32-37; <https://doi.org/10.29188/2712-9217-2025-11-3-32-37>

## The use of AI tools to improve the quality of scientific publications through automated analysis of article preprints

<https://doi.org/10.29188/2712-9217-2025-11-3-32-37>

G.M. Zhigulin<sup>1,2</sup>

<sup>1</sup> Department of Digital Medicine, I.M. Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia

<sup>2</sup> Morozov Children's City Clinical Hospital, Moscow, Russia

**Contact:** G.M. Zhigulin, [glebzhigulin.pp@gmail.com](mailto:glebzhigulin.pp@gmail.com)

### Summary:

**Introduction.** Amid the exponential growth of medical knowledge and publication activity, the speed of disseminating scientific results to the professional community is becoming a critical factor. An assessment of the duration of the peer-review and publication process in biomedical journals indicates that significant delays are often caused not by a lack of scientific novelty, but by formal non-compliance and technical errors in manuscripts.

**Objective.** The development and substantiation of the implementation of an automated artificial intelligence (AI)-based system for the preliminary analysis of preprints (the Ptolemaea project), aimed at reducing the rate of formal errors and accelerating the publication cycle.

**Materials and Methods.** The study is based on the principles of Data-Centric AI, scientometrics, and computational linguistics (NLP). The NLP model employed is used not for text generation, but for deep structural and stylistic analysis: detecting violations of IMRAD logic, formatting inconsistencies, verifying statistical indicators, and ensuring compliance with the criteria of specific journals.

**Results.** An analysis of publication activity shows that the average time from submission to acceptance ranges from 50 to 276 days. Notably, up to 93.2% of errors leading to revisions or rejections originate from the authors (formatting, statistics, incomplete data). The presented technology enables authors to identify these shortcomings at the preprint stage.

**Conclusion.** Automated preprint analysis is an effective practical tool that lowers the barrier to entry into scientific activity for young scientists and eliminates the routinization of the manuscript preparation process for experienced physician-researchers. The implementation of such systems contributes to improving the quality of scientific communication in healthcare.

**Key words:** artificial intelligence; scientific publications; preprints; NLP; peer review; medical statistics; Ptolemaea; scientometrics.

**For citation:** Zhigulin G.M. The use of AI tools to improve the quality of scientific publications through automated analysis of article preprints. Russian Journal of Telemedicine and E-Health 2025;11(3):32-37; <https://doi.org/10.29188/2712-9217-2025-11-3-32-37>

## ■ ВВЕДЕНИЕ

Современная биомедицинская наука характеризуется беспрецедентным ростом объема данных. Врачи и исследователи сталкиваются с необходимостью не только постоянно анализировать новую информацию, но и оперативно публиковать результаты собственных клинических наблюдений и экспериментальных работ. Однако существующая система научной коммуникации испытывает перегрузки. Традиционный процесс рецензирования (peer-review), являясь золотым стандартом контроля качества, часто становится «узким горлышком», задерживающим распространение критически важных медицинских знаний.

Проблема длительности публикационного цикла стоит особенно остро. По данным систематических обзоров, средний срок от подачи статьи до ее принятия (acceptance) в биомеди-

цинских журналах варьируется в широком диапазоне – от 50 до 276 дней, при этом медианные значения составляют от 21 до 248 дней [1]. Период от начала до завершения рецензирования занимает от 10 до 75 дней (медиана 29–87 дней), а коммуникация между рецензентом и автором по внесению правок требует еще от 24 до 73 дней [1, 2].

В контексте глобальных вызовов, таких как пандемия COVID-19, научное сообщество продемонстрировало способность ускорять процессы: время «от подачи до публикации» сократилось на 49% по сравнению с доковидным периодом для статей по соответствующей тематике [3]. Однако для работ по другим специальностям (кардиология, онкология, телемедицина) сроки остаются значительными.

Ключевой гипотезой данного исследования является предположение, что значительная часть задержек вызвана не отсутствием научной ►►

ценности работ, а формальными недостатками рукописей, которые могут быть устранены автоматически еще до этапа подачи в журнал.

*Целью работы* является описание методологии и функционала AI-инструмента (на примере разрабатываемой системы Ptolemaea), предназначенного для автоматизированного аудита научных препринтов.

### **Проблематика: анатомия публикационных задержек и ошибок**

Анализ причин возврата рукописей авторам или их отклонения (rejection) выявляет доминирование технических и методологических недочетов над концептуальными. Согласно библиометрическим исследованиям, до 93,2% ошибок в рукописях исходят непосредственно от авторов [4]. Спектр этих ошибок варьируется от банальных опечаток до серьезных методологических нарушений.

К наиболее частотным проблемам, препятствующим быстрой публикации, относятся:

1. Несоответствие требованиям конкретного журнала (Guide for Authors), отсутствие обязательных разделов, неправильное цитирование.

2. Неверный дизайн исследования, недостаточный размер выборки, ошибки в расчете р-значений (p-hacking или технические ошибки вычислений), неясное описание методов [4, 5].

3. Расхождения между данными в абстракте и основном тексте статьи, что затрудняет первичный скрининг редактором.

4. Грамматические и стилистические ошибки, особенно актуальные для авторов, публикующихся на неродном (чаще всего английском) языке.

Факторы, влияющие на скорость публикации, многогранны. Исследования показывают корреляцию между скоростью принятия статьи и такими параметрами, как конфликт интересов (члены редакции публикуются быстрее), география авторов (авторы из развитых стран имеют преимущество) и формат публикации (Open Access часто быстрее традиционной модели) [6, 7]. Тем не менее, качество подготовки рукописи остается единственным фактором, на который автор может повлиять напрямую и гарантированно.

## ■ МАТЕРИАЛЫ И МЕТОДЫ

В основу предлагаемого решения положена научная база Data-Centric AI, наукометрии и компьютерной лингвистики. Проект, получивший название Ptolemaea, реализуется на базе Цифровой кафедры Первого МГМУ им. И.М. Сеченова [8].

### *Технологический стек и алгоритмы*

В отличие от генеративных моделей (таких как GPT), задача которых – создание нового контента, в данном исследовании применяются NLP-модели (Natural Language Processing) для глубокого аналитического разбора текста. Подход можно охарактеризовать как «экспертная система на базе ИИ».

Процесс автоматизированного анализа препринта включает следующие этапы:

1. Парсинг и структурирование. Алгоритм декомпозирует загруженный файл (docx, pdf) на структурные элементы согласно стандарту IMRAD (Introduction, Methods, Results, and Discussion).

2. Формальная верификация. Проверка наличия всех обязательных разделов, соответствия объема аннотации, корректности оформления таблиц и рисунков, наличия ссылок на эти элементы в тексте.

3. Стилистический и лингвистический анализ. Выявление грамматических ошибок, стилистических несоответствий научному стилю, проверка уникальности текста (интеграция с базами научных статей).

4. Статистический аудит. Это наиболее инновационный модуль системы. Алгоритм сканирует раздел «Результаты» на предмет упоминания статистических критериев и р-значений, проверяя их внутреннюю согласованность и математическую возможность заявленных результатов при указанных размерах выборки.

5. Проверка библиографии. Верификация списка литературы, проверка соответствия ссылок внутри текста списку литературы, а также валидация DOI.

### **Валидация**

Эффективность разрабатываемой модели оценивается путем сравнения результатов авто-

матического анализа с рецензиями экспертов-людей. Тестовая выборка включает препринты, размещенные в открытых репозиториях (arXiv, bioRxiv, medRxiv), что позволяет обучать модель на реальных данных, содержащих типичные ошибки [9].

Внедрение системы автоматизированного анализа препринтов позволяет трансформировать процесс подготовки статьи. Технология помогает авторам заранее, до подачи в редакцию, увидеть и исправить ошибки, экономя время и повышая шансы на успешную публикацию (acceptance rate).

Сравнительный анализ функционала разрабатываемой системы Ptolemaea с существующими зарубежными аналогами (Penelope.ai (Великобритания), Stat Reviewer (США/Aries System), Manuscript Manager (США)) и отечественными решениями (НейроАссистент научного издательства, НЭИКОН) демонстрирует конкурентные преимущества предлагаемого подхода (табл. 1).

## ■ РЕЗУЛЬТАТЫ

Внедрение системы автоматизированного анализа препринтов позволяет трансформировать процесс подготовки статьи. Технология помогает авторам заранее, до подачи в редакцию, увидеть и исправить ошибки, экономя время и повышая шансы на успешную публикацию (acceptance rate).

Сравнительный анализ функционала разрабатываемой системы Ptolemaea с существующими

зарубежными аналогами (Penelope.ai (Великобритания), Stat Reviewer (США/Aries System), Manuscript Manager (США)) и отечественными решениями (НейроАссистент научного издательства, НЭИКОН) демонстрирует конкурентные преимущества предлагаемого подхода (табл. 1).

Ключевым отличием системы Ptolemaea является комплексный подход, включающий проверку статистической корректности, что реализовано лишь в единичных зарубежных продуктах (Stat Reviewer), доступ к которым для российских исследователей может быть ограничен.

Для образовательных и научных учреждений внедрение подобных систем несет прямой экономический эффект. По данным статистического сборника «Индикаторы науки: 2025», в России насчитывается более 800 институтов и университетов и 338 тысяч авторов, а финансирование науки превышает 1,6 трлн рублей [10]. При этом до 30% отчетов по грантам требуют существенной доработки, что влечет административные и финансовые издержки.

Внедрение AI-анализа препринтов позволяет:

- Для ВУЗов: Повысить показатели публикационной активности (KPI), улучшить качество обучения студентов и аспирантов академическому письму, обеспечить мониторинг качества исследований.
- Для молодых ученых: Снизить порог вхождения в публикационную деятельность. Система выступает в роли ментора, указывая на ошибки, характерные для начинающих исследователей. ►►

**Таблица 1. Сравнительный анализ функционала систем автоматической проверки научных статей**  
**Table 1. Comparative analysis of the functionality of automatic verification systems for scientific articles**

| Функция  | Ptolemaea (РФ) | Penelope.ai (UK) | Stat Reviewer (USA) | Manuscript Manager (USA) | НейроАссистент (РФ) |
|--|----------------|------------------|---------------------|--------------------------|---------------------|
| Оценка формальных параметров (структура, объем, рисунки) | +              | +                | -                   | +                        | +                   |
| Оценка статистической части (методы, p-value)            | +              | -                | +                   | -                        | -                   |
| Система рекомендаций по улучшению                        | +              | -                | -                   | -                        | +                   |
| Подбор журнала для публикации                            | +              | -                | -                   | +                        | +                   |
| Проверка на антиплагиат и цитирование                    | +              | +                | -                   | +                        | +                   |

- Для опытных авторов: Решить проблему рутинизации, делегируя ИИ проверку форматирования и библиографии, фокусируясь на научной новизне.

- Для издательств: Обеспечить предварительную фильтрацию входящего потока статей, снижая нагрузку на рецензентов и редакторов.

Потенциальный рынок (SAM) в России оценивается в 875 млн рублей в год с возможностью достижения выручки более 80 млн рублей к третьему году реализации проекта при охвате около 10% рынка.

## ■ ОБСУЖДЕНИЕ

Традиционная модель закрытого рецензирования подвергается критике из-за длительности, предвзятости и неспособности выявить все ошибки. В ответ на это возникают новые модели:

1. Fast-track. Ускоренное рецензирование (часто платное) [11].

2. Publish–Review–Curate. Модель, где статья сначала публикуется как препринт, затем проходит рецензирование и кураторскую оценку [12].

3. Открытый обзор. Публикация комментариев рецензентов вместе с текстом статьи.

Использование AI-инструментов гармонично вписывается в парадигму Publish–Review–Curate и развития препринт-серверов. Активность российских авторов на платформе arXiv в 2024 году выросла на 12%, что подтверждает готовность сообщества к новым форматам коммуникации. AI-валидация выступает первичным фильтром качества, гарантирующим, что препринт соответствует базовым стандартам научности перед тем, как он попадет к экспертам-людям.

Важно подчеркнуть, что AI-инструменты, такие как Ptolemaea, не заменяют научного рецензента. Искусственный интеллект на текущем этапе развития не способен оценить истинную клиническую значимость, новизну идеи или этич-

ность проведенного эксперимента над пациентами. Задача AI – взять на себя функцию «технического редактора» и «статистического контролера».

Для врача-клинициста использование такого инструмента означает возможность сосредоточиться на медицине, а не на борьбе с требованиями к оформлению списка литературы. Это практический инструмент, снижающий риск отказа по формальным признакам, который особенно демотивирует авторов.

## ■ ЗАКЛЮЧЕНИЕ

Автоматизированный анализ препринтов с использованием технологий Data-centric AI и NLP представляет собой перспективное направление в области медицинской информатики и наукометрии. Проект Ptolemaea демонстрирует возможность создания комплексного ответственного решения, превосходящего по ряду параметров зарубежные аналоги, в частности, за счет модуля статистического аудита.

Внедрение данной технологии позволит:

1. Сократить время от написания до публикации статьи за счет устранения формальных ошибок на раннем этапе (shift-left testing).

2. Снизить нагрузку на редакции научных журналов и рецензентов.

3. Повысить общее качество биомедицинских публикаций, минимизируя количество статей с некорректным дизайном или статистикой.

4. Стимулировать публикационную активность молодых ученых и студенческих сообществ, которых в России насчитывается более 500.

Дальнейшее развитие проекта предполагает расширение датасетов для обучения моделей, интеграцию с редакционными системами российских журналов и совершенствование алгоритмов проверки семантической связности текста. //

## ЛИТЕРАТУРА

1. Andersen MZ, Fonnes S, Rosenberg J. Time from submission to publication varied widely for biomedical journals: a systematic review. *Curr Med Res Opin* 2021;37(6):985–993. <https://doi.org/10.1080/03007995.2021.1905622>
2. Maggio LA, Bynum WE 4th, Schreiber-Gregory DN, Durning SJ, Artino AR Jr. When will I get my paper back? A replication study of

- publication timelines for health professions education research. *Perspect Med Educ* 2020;9:253–256.
3. Horbach SPJM. Pandemic publishing: Medical journals strongly speed up their publication process for COVID-19. *Quant Sci Stud* 2020;1(3):1056–1067. [https://doi.org/10.1162/qss\\_a\\_00076](https://doi.org/10.1162/qss_a_00076)
4. Soleimanpour S, Sedghi S, Asghari H, Nemati-Anaraki L. No study

## ЛИТЕРАТУРА

is ever flawless: A scoping review of common errors in biomedical manuscripts. *Account Res* 2021;29(6):397–414. <https://doi.org/10.1080/08989621.2021.1937604>

5. Vural S, Kaya H, Coşkun F. A bibliometric study on the publication errors in emergency medicine journals from 2000 to 2020. *Am J Emerg Med* 2022;60:140–144. <https://doi.org/10.1016/j.ajem.2022.08.001>

6. Sebo P, Fournier JP, Ragot C, Gorioux P, Herrmann F, Maisonneuve H. Factors associated with publication speed in general medical journals: a retrospective study of bibliometric data. *Scientometrics* 2019;119:1037–1058. <https://doi.org/10.1007/s11192-019-03061-8>

7. Taşkın Z, Taşkın A, Doğan G, et al. Factors affecting time to publication in information science. *Scientometrics* 2022;127:7499–7515. <https://doi.org/10.1007/s11192-022-04296-8>

8. Жигулин Г.М., и др. Использование AI инструментов для повышения качества научных публикаций. Материалы конференции ИТМ. 2024.

9. Accelerating scientific progress with preprints. *Nat Comput Sci* 2024;4:311. <https://doi.org/10.1038/s43588-024-00641-4>

10. Гохберг Л.М., Дитковский К.А., Евневич Е.И. и др. Индикаторы науки: 2025: статистический сборник. М.: НИУ ВШЭ, 2025.

11. Teixeira da Silva JA, Yamada Y. Accelerated Peer Review and Paper Processing Models in Academic Publishing. *Pub Res Q* 2022;38:599–611. <https://doi.org/10.1007/s12109-022-09891-4>

12. Eisen MB, Akhmanova A, Behrens TE, et al. Scientific Publishing: Peer review without gatekeeping. *eLife* 2022;11:e83889.

## Сведения об авторе:

Жигулин Г.М. – аспирант ГБУЗ «Морозовская ДГКБ ДЗМ», сотрудник кафедры цифровой медицины Первого МГМУ им. И.М. Сеченова, руководитель проекта Ptolemaea, Москва, Россия

## Вклад автора:

Жигулин Г.М. – определение научного интереса, дизайн и технологическое решение проекта, обзор литературы, написание текста, общее руководство проектом, 100%

**Конфликт интересов:** Автор заявляет о разработке описываемого программного продукта Ptolemaea.

**Финансирование:** Исследование проведено без спонсорской поддержки.

**Статья поступила:** 09.05.25

**Рецензирование:** 31.05.25

**Принята к публикации:** 17.07.25

## Information about author:

Zhigulin G.M. – postgraduate student at the Morozov Children's City Clinical Hospital of the Moscow Health Department, employee of the Department of Digital Medicine at the I.M. Sechenov First Moscow State Medical University, and head of the Ptolemaea project, Moscow, Russia

## Author Contribution:

Zhigulin G.M. – definition of scientific interest, design and technological solution of the project, literature review, writing, overall project management, 100%

**Conflict of interest.** The author declares the development of the described software product Ptolemaea.

**Financing.** The study was performed without external funding

**Received:** 09.05.25

**Review:** 31.05.25

**Accepted for publication:** 17.07.25