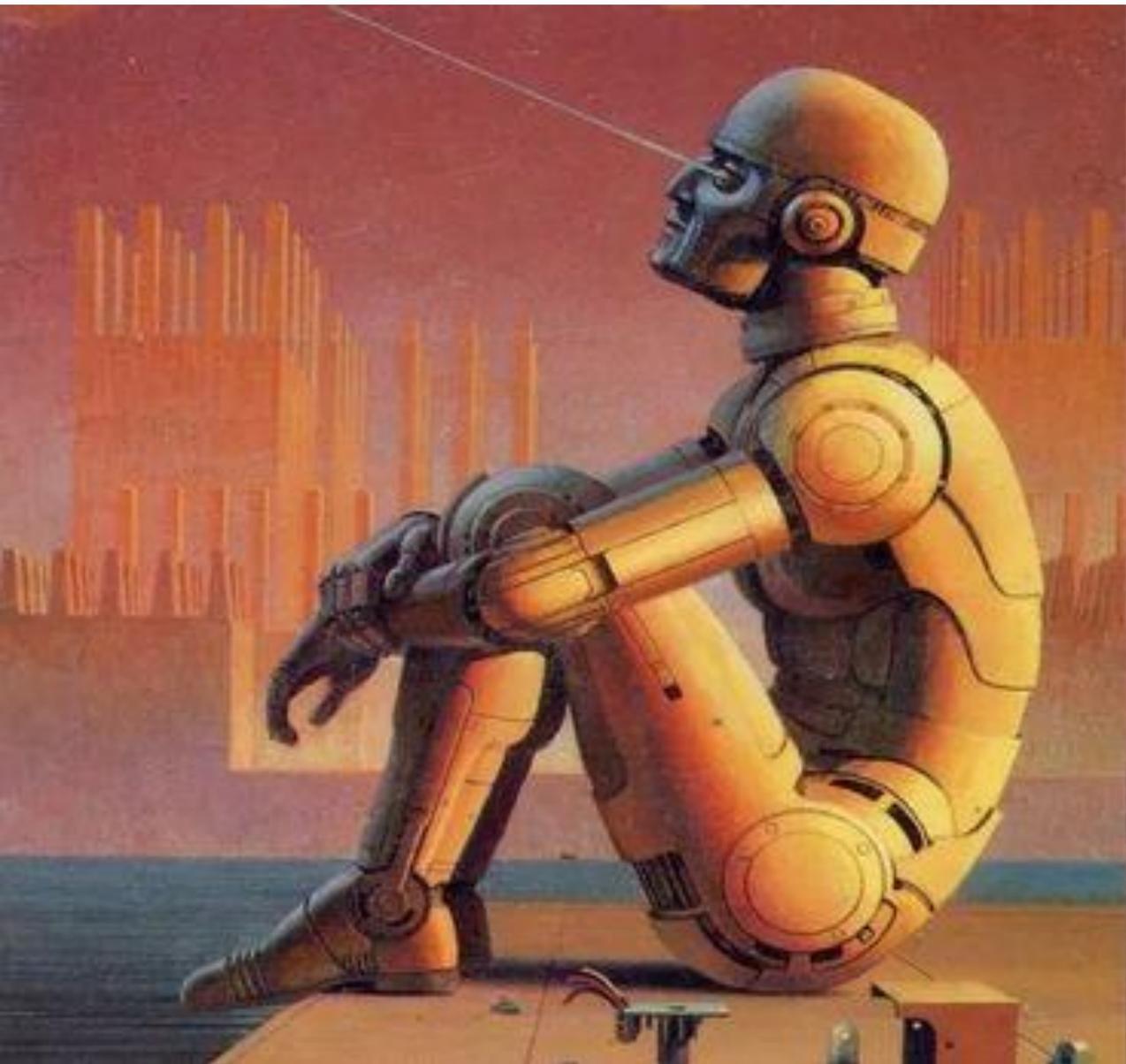


# Детектор ИИ и детектор дубликатов. О некоторых новинках системы «Антиплагиат»

---

Юрий Викторович Чехович, к.ф.-м.н.  
Президент компании Антиплагиат



# Детектор ИИ

---

# Примеры задач, которые решают с помощью ИИ

## Тексты

Определение авторства

Выявление ошибок

Переводные заимствования

Машинная генерация текста

...

## Изображения

Машинная генерация изображений

Выделение изображений

Заимствования изображений и графиков

Рукописные заимствования

...

## Прочее

Структура документа

Формулы

Таблицы

Генерация поисковых запросов

...

# Бум генеративных сервисов



YandexGPT

17 мая 2023 г.



Writesonic  
2020 г.



Rytr

Июнь 2023 г.



GIGA  
CHAT

Апрель 2023 г.



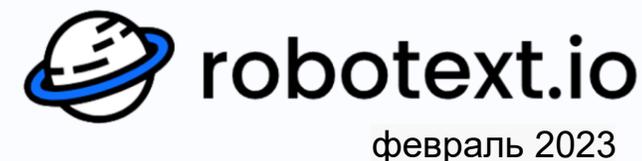
Gerwin  
2022 г.



YOU.com  
9 ноября 2021 г.



perplexity  
Август 2022 г.



robotext.io  
февраль 2023



CHATSONIC  
4 марта 2023 г.



character.ai  
16 сентября 2022 г.

# Почему генеративные сервисы востребованы?

---

- Доступность – диалоговый режим работы создает ощущение, что общаешься с живым человеком
- Скорость – можно быстро получить много текста
- Натуральность – текст очень похож на «человеческий»
- Итеративность – результат можно «подгонять под себя»
- Оригинальность (формальная) – полученный текст отсутствует в каких-либо базах и оценивается детекторами плагиата как оригинальный

# Высокая оригинальность сгенерированного текста

ГЛАВНАЯ / КАБИНЕТ

## Кабинет

ПОИСК ПО НАЗВАНИЯМ ДОКУМЕНТОВ

УДАЛЕННЫЕ ДОКУМЕНТЫ 1/1

ПЕРЕПРОВЕРИТЬ ПЕРЕМЕСТИТЬ УДАЛИТЬ ИСТОРИЯ ОТЧЕТОВ В ИЗ ИНДЕКСА

Название	Дата загрузки	Оригинальность	
<input type="checkbox"/> DOCX sample-1	16 Мая 2023 13:35	97,62%	<a href="#">ПОСМОТРЕТЬ РЕЗУЛЬТАТЫ</a>
<input type="checkbox"/> DOCX sample-2	16 Мая 2023 13:35	98,16%	<a href="#">ПОСМОТРЕТЬ РЕЗУЛЬТАТЫ</a>

ПАПКИ: Все документы

- Корневая папка 46
- Кейс Совета по этике 4
- Яненко 12
- Кейсы GPT 2**

ПРОВЕРЕНО: 16.05.2023 10:42:10

1,2%	СОВПАДЕНИЯ
0%	САМОЦИТИРОВАНИЯ
1,17%	ЦИТИРОВАНИЯ
97,63%	ОРИГИНАЛЬНОСТЬ

ПРОВЕРЕНО: 16.05.2023 10:40:27

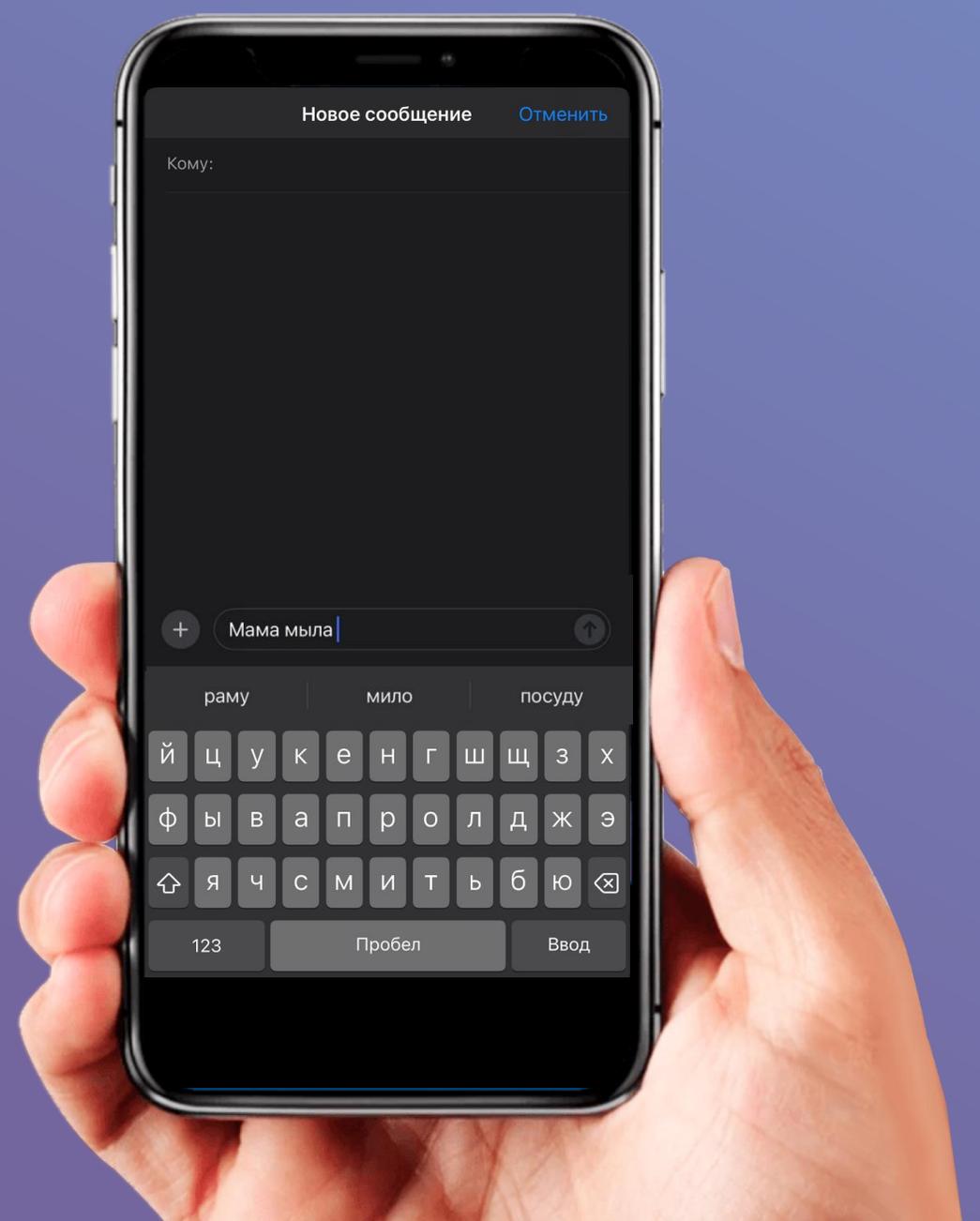
0,83%	СОВПАДЕНИЯ
0%	САМОЦИТИРОВАНИЯ
1,01%	ЦИТИРОВАНИЯ
98,16%	ОРИГИНАЛЬНОСТЬ

# Недостатки генеративных моделей

---

- «Галлюцинации» – недостоверные сведения
- Бедность (предсказуемость) стилистики
- Логические противоречия
- Повторы
- Неадекватные ссылки

# T9



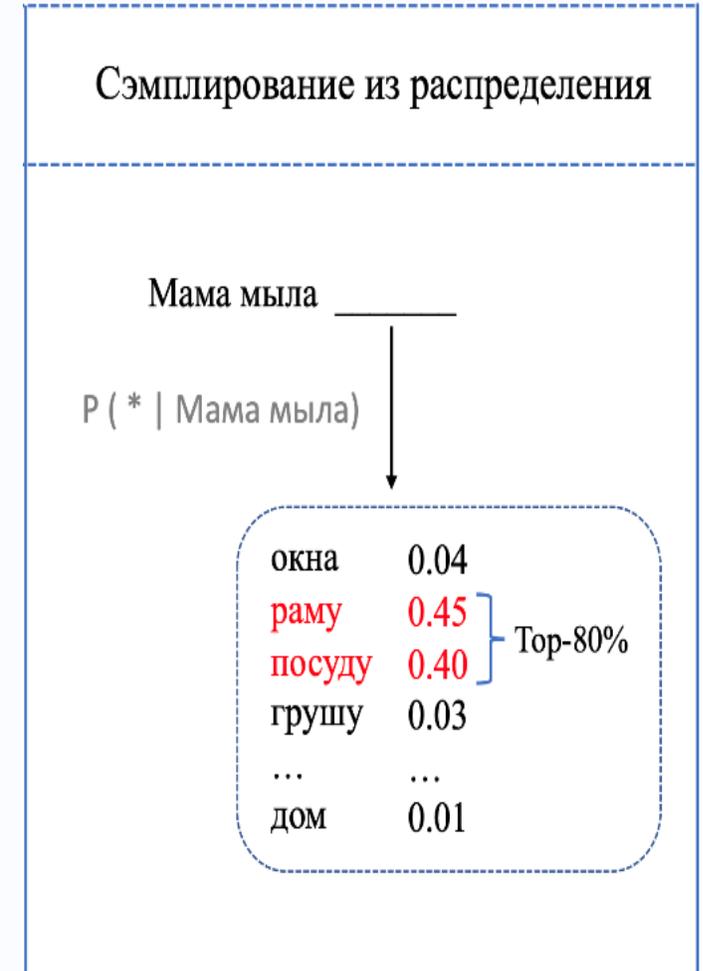
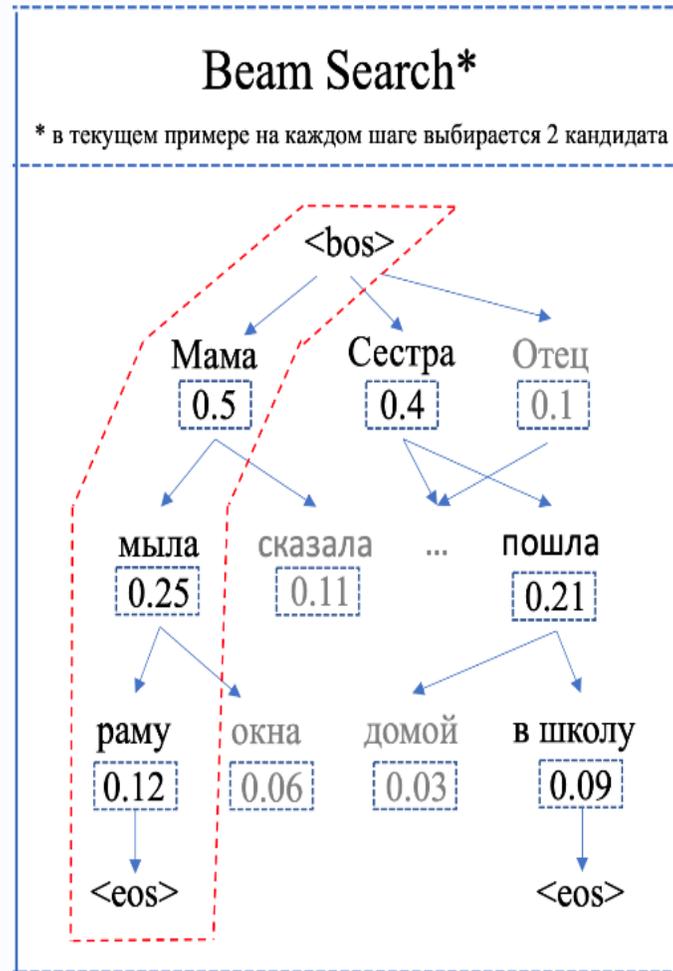
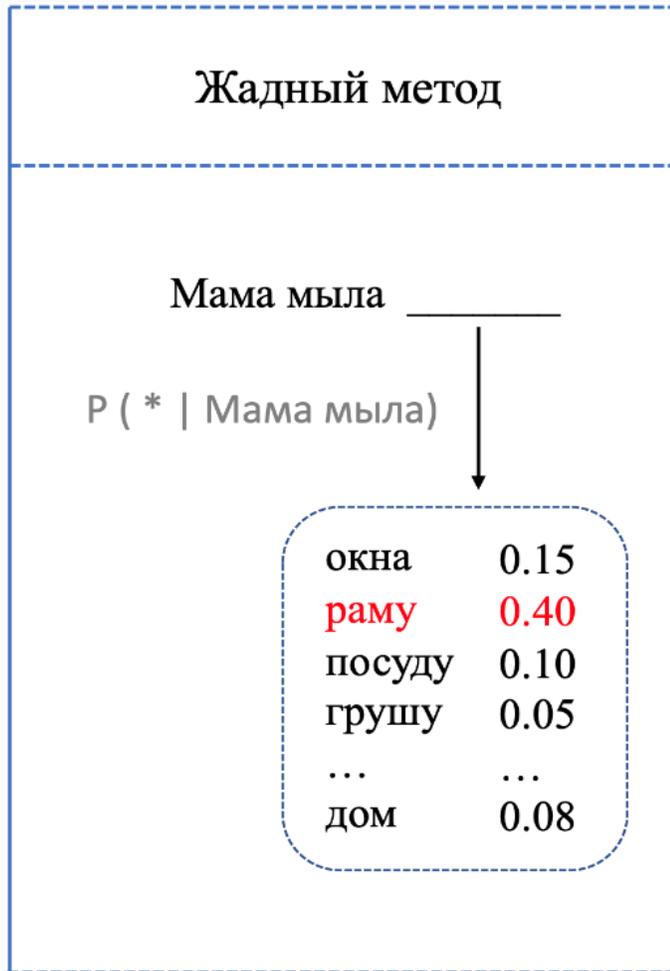
# Как работают генеративные сервисы

Нужно проанализировать очень много текстов



# Генеративная модель – алгоритм продолжения текста

## Мама мыла... [как продолжить?]



# Рецепт генеративного ИИ

---

Для генерации текста используется **вероятностный алгоритм**, обученный на огромном количестве текстов и дообученный на данных пользователей, принципиально **не способный критически оценивать качество и достоверность полученного результата.**

# Модели генеративного искусственного интеллекта с полным их разоблачением



Ю. В. Чехович, А. Грабовой, Г. Грицай // Университетская книга. – 2024. – № 5. – С. 58-65. – EDN YXSKBC.

<https://www.unkniga.ru/ai/modeli-generativnogo-iskusstvennogo-intellekta-s-polnym-ikh-razoblacheniem.html>

# Обнаружение сгенерированного текста

11 мая 2023

## Антиплагиат выявляет текст, созданный ChatGPT



В системе обнаружения текстовых заимствований «Антиплагиат» появился новый функционал – автоматическое выявление машинно-сгенерированных текстов. Новый функционал будет доступен всем пользователям платных версий системы с мая 2023 года.

<https://antiplagiat.ru/news/ext-chatgpt/>

# Выявление сгенерированного текста в системе Антиплагиат

## Схема алгоритма



Г. М. Грицай, А. В. Грабовой, А. С. Кильдяков, Ю. В. Чехович Поиск искусственно сгенерированных текстовых фрагментов в научных документах // Доклады Российской академии наук. Математика, информатика, процессы управления. – 2023. – Т. 514, № 2. – С. 308-317. – DOI 10.31857/S2686954323601677. – EDN GQRWLF.

# Детектор сгенерированного текста в системе Антиплагиат

The screenshot displays the 'Подозрительный документ' (Suspicious document) page in the AntiPlagiat system. The interface includes a header with the system logo, navigation links, and a breadcrumb trail: 'ГЛАВНАЯ / КАБИНЕТ / РЕЗУЛЬТАТЫ ПРОВЕРКИ / ПОДОЗРИТЕЛЬНЫЙ ДОКУМЕНТ'. The main content area shows a document preview with a page indicator '1/84'. On the right, a sidebar contains a warning 'ОТМЕТКА ВКЛЮЧЕНА' (Note included) and a summary of detected issues. A red box highlights the 'СГЕНЕРИРОВАННЫЙ ТЕКСТ' (Generated text) category, which includes a tooltip explaining that it contains fragments likely created by artificial intelligence.

**АНТИПЛАГИАТ**  
ОБНАРУЖЕНИЕ ЗАИМСТВОВАНИЙ

Sk Skolkovo  
Реестр отечественного ПО

КРАТКОЕ НАИМЕНОВАНИЕ

МЕНЮ RU

ГЛАВНАЯ / КАБИНЕТ / РЕЗУЛЬТАТЫ ПРОВЕРКИ / ПОДОЗРИТЕЛЬНЫЙ ДОКУМЕНТ

## Подозрительный документ

1/84

**ОТМЕТКА ВКЛЮЧЕНА**  ВКЛ

На этой странице отображены возможные попытки маскировки заимствований. Будьте внимательны при работе с документом

**В ДОКУМЕНТЕ** **НА СТРАНИЦЕ**

В документе обнаружена маскировка заимствований из следующих групп:

- ВСТАВКА** ⓘ В данную группу входят фрагменты, которые, предположительно, созданы с помощью искусственного интеллекта
- СГЕНЕРИРОВАННЫЙ ТЕКСТ** ⓘ ▾

# Детектор сгенерированного текста в системе Антиплагиат

↑ НАВЕРХ

5/50

В ДОКУМЕНТЕ НА СТРАНИЦЕ

ПОКАЗАТЬ ВСЕ СКРЫТЬ ВСЕ

СГЕНЕРИРОВАННЫЙ ТЕКСТ ⓘ

Искусственный текст (beta)

Например, на основе финансового анализа можно определить наиболее эффективные способы использования свободных денежных средств компании, принять решение о необходимости привлечения дополнительных финансовых ресурсов, выбрать наиболее оптимальный тип кредитования или выявить причины снижения прибыли компании. [6].

Важно отметить, что финансовый анализ не является единственным инструментом для оценки эффективности деятельности компании. Кроме того, необходимо учитывать многие другие факторы, такие как рыночная конъюнктура, конкуренция, риски и т.д. Однако финансовый анализ является важным компонентом в общей системе управления финансами компании.

Таким образом, сущность и значение финансового анализа заключается в оценке финансового состояния компании, выявлении проблемных зон и определении направлений для дальнейшего развития. Основными задачами финансового анализа являются анализ финансовой устойчивости, рентабельности, ликвидности, дебиторской и кредиторской задолженности, управления активами и капиталом, а также анализ инвестиционной деятельности.

Для проведения финансового анализа необходимо иметь знания и умения в области финансовой отчетности и финансовых показателей, а также опыт в их интерпретации и применении. [7].

Для проведения финансового анализа используются различные методы и инструменты, такие как горизонтальный и вертикальный анализ финансовой отчетности, показатели анализа рентабельности, ликвидности и финансовой устойчивости, денежный поток, анализ дебиторской и кредиторской задолженности, а также сравнение финансовых показателей с аналогичными компаниями на рынке.

Однако, при использовании финансового анализа необходимо учитывать его ограничения. Во-первых, финансовый анализ основывается на финансовой отчетности, которая может содержать ошибки, неполные данные или быть подвержена манипуляциям со стороны руководства компании. Во-вторых, финансовый анализ не учитывает факторы, связанные с внешней средой

# Сессия 2024 года: каждая пятая работа российских студентов имеет следы ИИ

Каждая пятая работа российских студентов имеет следы ИИ.



АНТИПЛАГИАТ



Юрий  
Чехович



Андрей  
Грабовой



[https://antiplagiat.ru/news/ai\\_in\\_papers2024/](https://antiplagiat.ru/news/ai_in_papers2024/)

# Признаки использования ИИ

- Несогласованность, нарушения связности
- Много банальностей («эффект жвачки»)
- Фразы, не несущих смысловой нагрузки
- Повторы («эффект рыбки Дори»)
- Нарушения логики построения текста
- Нет ссылок там, где они необходимы
- Ссылки в оригинальном тексте
- Стилистическая неоднородность

141

организациями по всему миру, способствовала построению нового типа международных отношений путем построения нового типа отношений политических партий, способствовала улучшению глобальной сети партнерства путем укрепления глобальной сети партнерства политических партий. По случаю столетия со дня основания Коммунистической партии Китая зарубежные политические партии и организации направили более 1500 поздравительных посланий. Были успешно проведены такие глобальные мероприятия политических партий, как Диалог на высоком уровне между Коммунистической партией Китая и мировыми политическими партиями и Саммит лидеров Коммунистической партии Китая и мировых политических партий. Генеральный секретарь Си Цзиньпин выступил с программной речью, которая определила векание времени и направление человеческого прогресса. Коммунистическая партия Китая и политические партии стран-партнеров «Один пояс один путь» создали механизм совместного обсуждения политических партий и продолжают формировать международный консенсус по продвижению высококачественного развития «Один пояс один путь». Китай также провел международные конференции гражданского общества для совместной реализации инициатив глобального развития, предложенных генеральным секретарем Си Цзиньпином, и отметил Международный день мира, содействуя реализации инициатив глобального развития и глобальной безопасности, а также сотрудничая с политическими партиями по всему миру для продвижения общих ценностей всего человечества и построения лучшего мира.

В этом десятилетии Китай активно продвигает построение сообщества человеческой судьбы. Основным содержанием сообщества человеческой судьбы является построение прочного мира, всеобщей безопасности, всеобщего процветания, открытости, всеохватности, чистоты и красоты. Китай является как теоретиком, так и практиком построения сообщества человеческой судьбы. Китай подписал более 200 документов о сотрудничестве со 149 странами и 32 международными организациями.

# Дублированные публикации

---



# Анализ дублированных публикаций в российских журналах



Journal of Informetrics  
Volume 16, Issue 1, February 2022, 101246



<https://doi.org/10.1016/j.joi.2021.101246>

## Analysis of duplicated publications in Russian journals

Yury V. Chekhovich <sup>1</sup>  , Andrey V. Khazov 



### Abstract

The article presents a study of publication ethics violations in Russian-language scientific publications related to the duplicated publication. The aim of the study is to assess the frequency of the above-noted violations in the data of eLIBRARY.RU, the largest aggregator of full texts of Russian-language scientific publications. For the purposes of the study, we used the tools of the “Antiplagiat” plagiarism detection system. Out of the almost 12 million full-text publications on the eLIBRARY.RU platform, more than 3.8 million scientific articles were selected for analysis. The study identified 70 406 cases of duplication of publications. In each of the detected cases, the same – or significantly similar – texts were published two or more (up to 73) times. An analysis of the most significant cases by number of publications is presented along with a detailed discussion of examples of the identified violations. A significant increase in the number and proportion of duplicated publications was identified in the period from 2014 to 2017. Conclusions are presented concerning shortcomings in editorial processes that allow duplication of publications along with the factual impossibility of detecting duplication in cases of simultaneous submission of the manuscript to different journals, the insufficient use by journal editors of means for detecting inappropriate borrowings and the need to retract a significant number of articles in connection with the identified violations.

# Некоторые выводы

---

- Было обнаружено 70 407 случаев дублирования публикаций
- Всего в кластеры дублированных статей попали 3,88 % всех исследованных публикаций
- Значительная часть случаев представляют собой примеры неэтичного дублирования; в настоящее время статьи не ретрагированы
- Действующий издательский процесс затрудняет журналам возможности своевременно обнаружить попытку самоплагиата
- Нужен сервис для отслеживания самоплагиата после публикации

# Результаты в открытом доступе

Все данные, полученные в описанном исследовании, находятся в открытом доступе

<http://dx.doi.org/10.17632/dy2smpf277.2>

Mendeley Data Find Research Data My Data ? YC

## Open dataset of duplicate publications

Published: 24 March 2021 | Version 2 | DOI: 10.17632/dy2smpf277.2  
Contributors: Yury Chekhovich, Andrey Khazov

### Description

Dataset for "Duplicate publication: Analysis of scientific misconduct in Russian journals" paper. Contains information about near-duplicate publications in Russian scientific library eLIBRARY.RU

[Download All 5 MB](#)

### Files

 clusters_final_susp.xlsx	5 MB <a href="#">↓</a>
---	------------------------

### Dataset metrics

Usage

Views:	251
Downloads:	30

 [View details >](#)

### Latest version

Version 2  
Published: 24 Mar 2021  
DOI: 10.17632/dy2smpf277.2

Cite this dataset

Chekhovich, Yury; Khazov, Andrey (2021), "Open dataset of duplicate publications", Mendeley Data, V2, doi: 10.17632/dy2smpf277.2

[Copy to clipboard](#)



# Данные (2024)

---

Исследование проводилось на данных НЭБ eLIBRARY.ru (<https://www.elibrary.ru/>)

Почти **23 млн** публикаций с полными текстами

Фильтрация:

- Тип и категория публикации:
  - тип «Научная статья»,
  - категории «Статья в журнале», «Краткое сообщение», «Обзорная статья», «Не определен»
- Язык публикации: русский
- Длина публикации: не менее 5000 знаков
- Полнота метаданных: есть информация об авторах, указан год публикации

# Данные и результаты

---

**2019**

Почти **12 млн** публикаций с полными текстами

После фильтрации среди анализируемых данных остается **3 803 964** публикации

**70 406** кластеров,  
включающих **147 708** статей

**2024**

Более **23 млн** публикаций с полными текстами

После фильтрации среди анализируемых данных остается **5 327 980** публикации

**195 267** кластеров,  
включающих **423 265** статей,  
в том числе **14 892** кластера со  
статьями, изданными после 2020 года

# Сделано из ранее запланированного

---

- Взаимодействовать с журналами по вопросам ретракции **Готово**
- Провести исследование на обновленных данных **Готово**
- Опубликовать результаты на русском языке **В процессе**
- Сервис предупреждения о дублировании **Детектор дубликатов**

# Детектор дубликатов

ТАРИФЫ Full XL  [ИЗМЕНИТЬ](#)ПРОВЕРКИ 2  152д [ПРОВЕРИТЬ ДОКУМЕНТ](#)АДМИНИСТРАТОР test@test.ru  [ВОЙТИ В КАБИНЕТ](#) МЕНЮ ru 

ГЛАВНАЯ / КАБИНЕТ / РЕЗУЛЬТАТЫ ПРОВЕРКИ / ПОДОЗРИТЕЛЬНЫЙ ДОКУМЕНТ

## Подозрительный документ

ИСХОДНОЕ

1 < 1/23 >

*[Blurred document content]*

 **ОТМЕ**

На этой странице вы можете увидеть попытки ма... внимательн...

**в докум**

В документе заимствован...

**ПРОВЕРКА В ДРУГОЙ ОРГАНИЗАЦИИ**  

**Этот или очень похожий документ ранее проверялся в другой организации**

- Организация:** МГСУ  
**Дата проверки:** 12.07.2021
- Организация:** КазНУ  
**Дата проверки:** 22.04.2024
- Организация:** СПбГУПТД  
**Дата проверки:** 11.10.2024

Этот или очень похожий документ проверялся ранее в других организациях. Если вы проверяете ВКР, диплом или магистерскую диссертацию, скорее всего работа написана не самостоятельно, несмотря на возможно высокий показатель оригинальности. Если вы проверяете статью, доклад, докторскую или кандидатскую диссертацию, сравните список организаций с аффилиациями авторов. Если организации совпадают, вполне возможно документ был проверен перед подачей. В противном случае документ с большой вероятностью не является написанным самостоятельно. В любом случае рекомендуем получить у автора разъяснение, почему работа проверялась в другой организации.

# Спасибо за внимание!

---

Юрий Викторович Чехович  
[chehovich@antiplagiat.ru](mailto:chehovich@antiplagiat.ru)