

КНИГА+: проекты цифровой трансформации.
Эволюция бизнес-моделей и сервисов

Интеллект-
невидимка.
Как выявить
использование
искусственного
интеллекта авторами

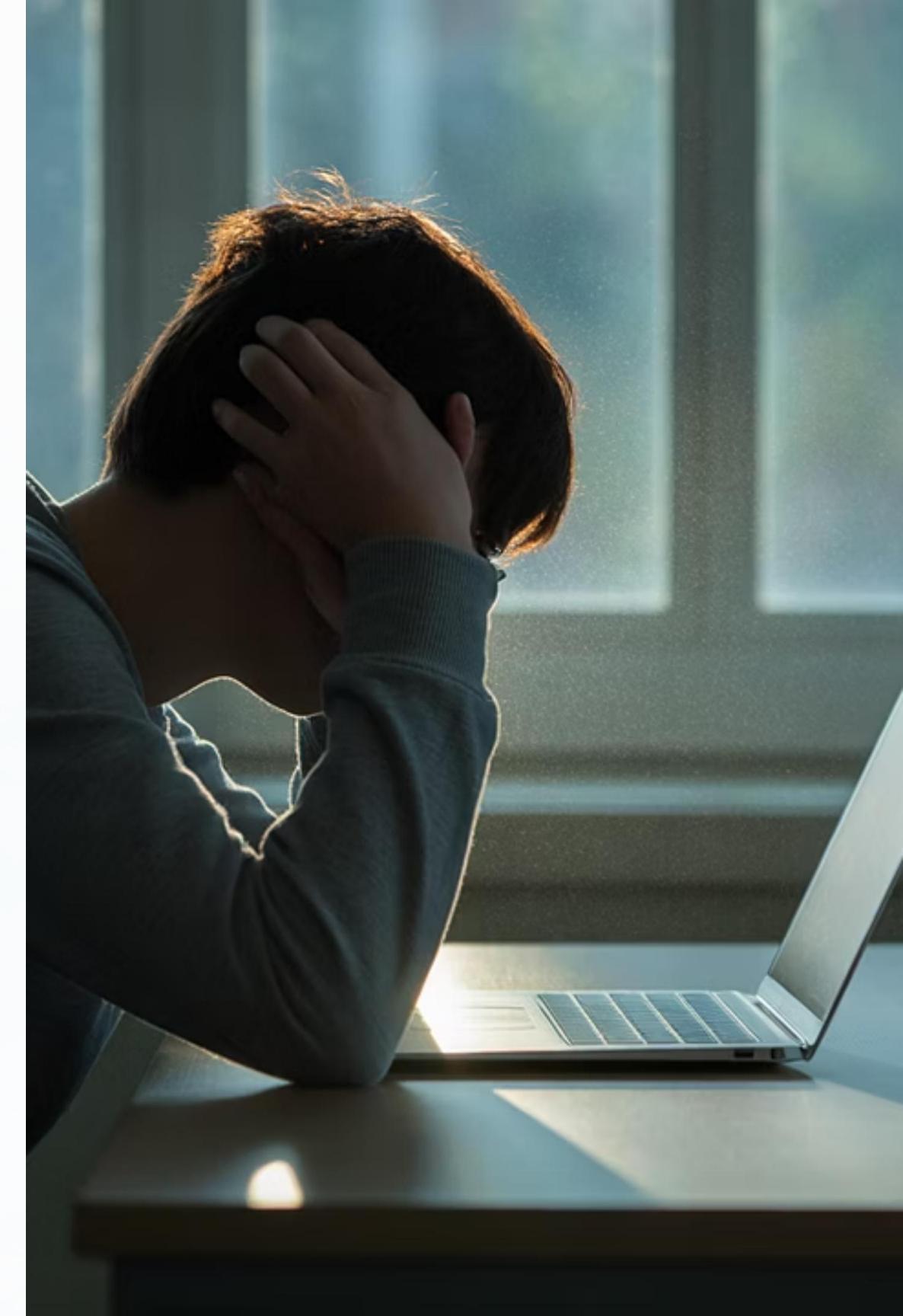
AI

Юрий Чехович

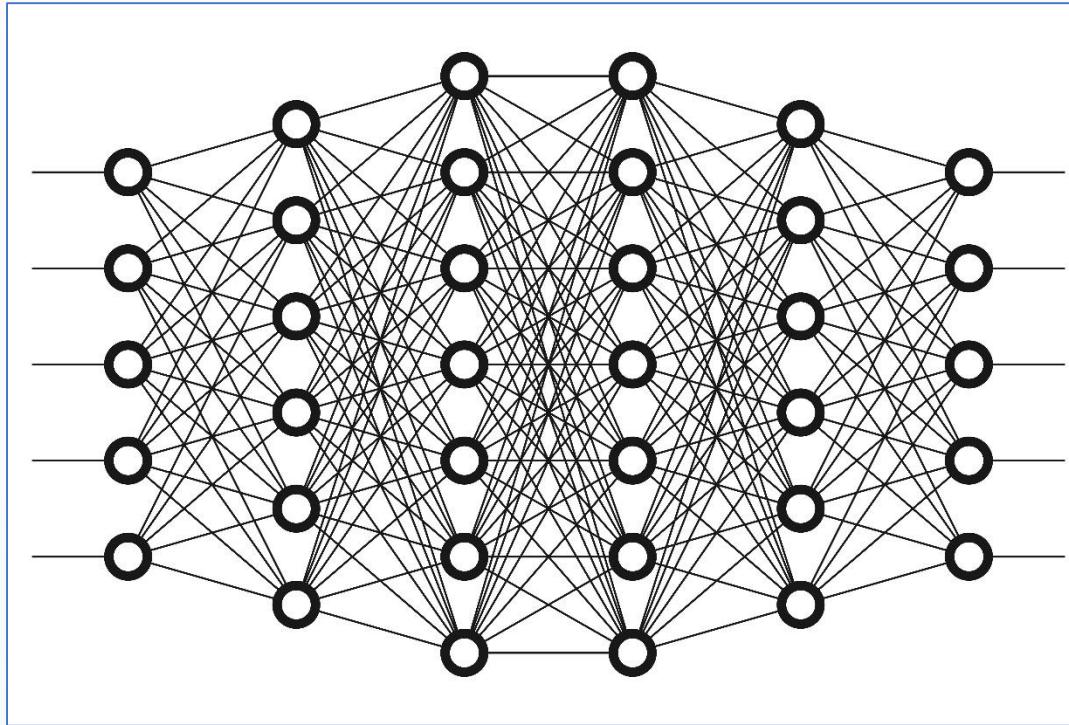
3 сентября 2025, ВДНХ, 57 павильон

Проблемы использования авторами ИИ

- Право и лицензирование данных для обучения моделей
- **Подделки и «книги-двойники» под именами реальных авторов**
- **Качество и безопасность контента**
- Операционные риски: массовые рассылки и их обнаруживаемость
- Отзывы и репутация
- Политики издательств и научных брендов
- Прозрачность, маркировка и происхождение контента



Как работает генеративный 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**.

Способы детекции ИИ-текста



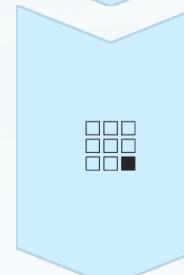
Алгоритмические методы

Статистический анализ текста, выявление паттернов, характерных для генеративных моделей.



Цифровые метки

Встраивание незаметных «водяных знаков» в тексты, генерируемые ИИ, для последующей идентификации источника.



Экспертная оценка

Выявление несоответствий в стиле, аргументации и логике текста опытными преподавателями и рецензентами.

Современные системы детекции используют комбинацию методов для повышения точности определения ИИ-авторства. Важным элементом проверки остается сопоставление текста с общим уровнем знаний и навыков студента, демонстрируемым в очных формах работы.

Grade

Text Papern Pectest irisq:

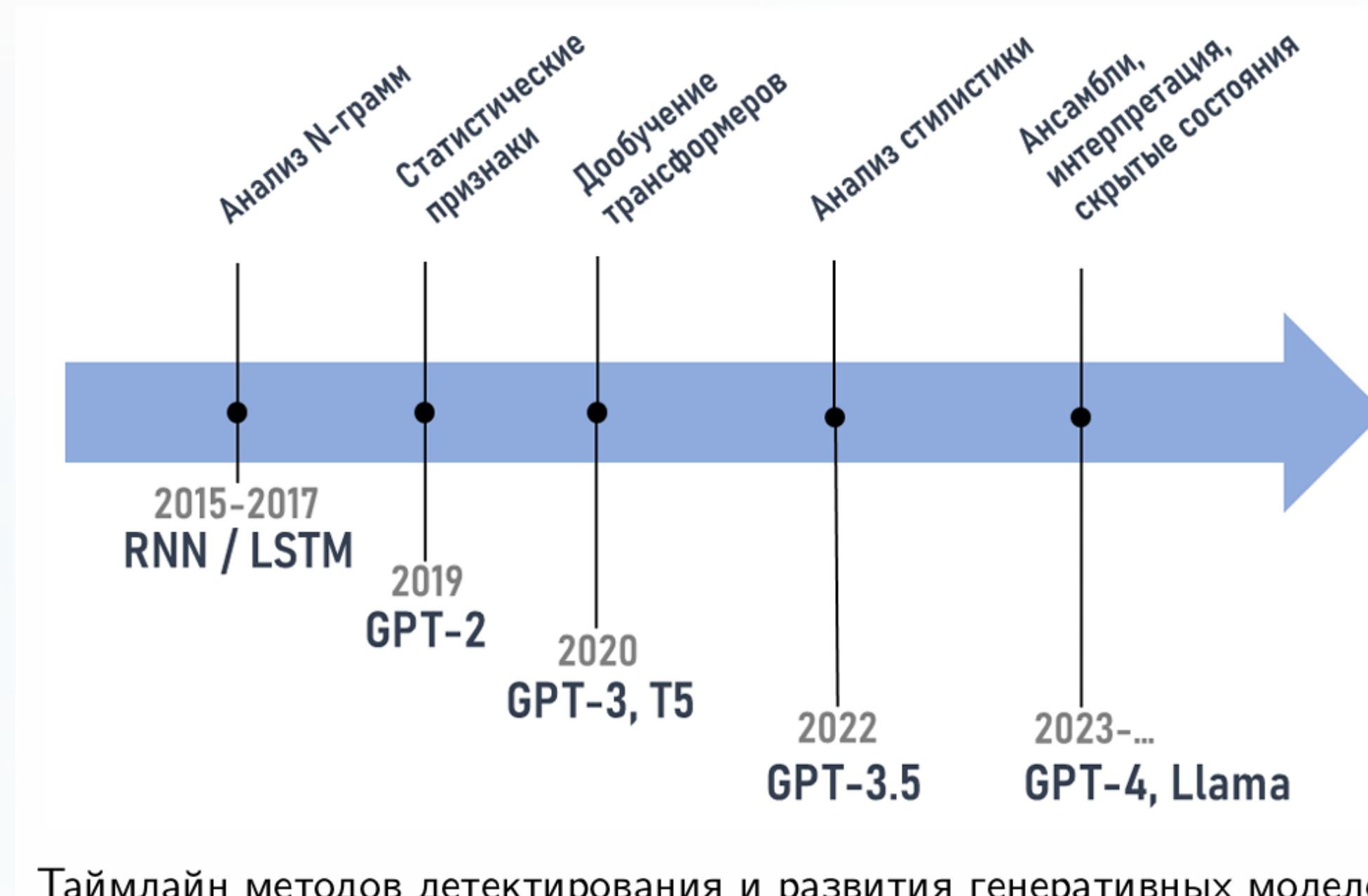
Keywords

Text

All Doctoral Certified Solved Recent

© 2024 EduAI Terms of Service Privacy Policy Accessibility

Алгоритмы детекции ИИ-текста



Проблемы, выявляемые в искусственном тексте

Признак	Объяснение
1. Синтаксические особенности	<ul style="list-style-type: none">• Длинные и сложные предложения• Сложная структура предложений
2. Лексические особенности	<ul style="list-style-type: none">• Техническая лексика/жаргон• Специализированная терминология• Формальный язык• Неформальный язык, сленг
3. Структурные особенности	<ul style="list-style-type: none">• Чёткая организация текста• Иерархическая структура• Подзаголовки, разделы• Маркированные списки
4. Особенности содержания	<ul style="list-style-type: none">• Конкретные детали/примеры• Наличие цитат• Исторический контекст• Абстрактные понятия
5. Стилистические особенности	<ul style="list-style-type: none">• Отсутствие связности• Логические нарушения• Стилистическая неоднородность• Сдвиг тона
6. Риторические особенности	<ul style="list-style-type: none">• Риторические вопросы• Переходные фразы• Сравнение/противопоставление
7. Другие особенности	<ul style="list-style-type: none">• Повторение фраз• Ссылки на источники



Проблемы систем детекции ИИ-текста

Технические ограничения

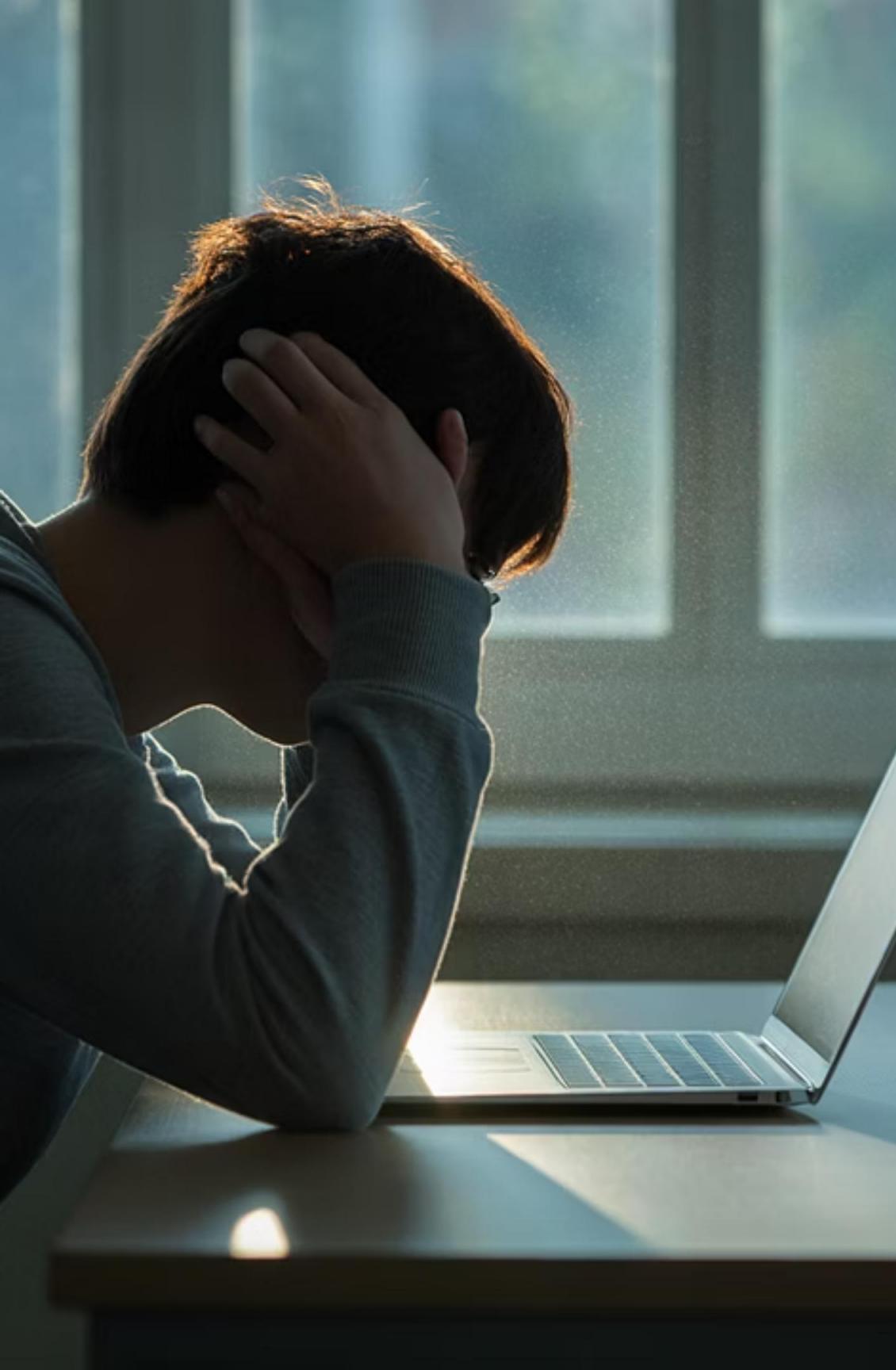
Высокий процент ложноположительных и ложноотрицательных результатов. Детекторы часто ошибочно маркируют как ИИ-сгенерированные тексты иностранных авторов или тексты на специфические технические темы.

Эволюция моделей и методов маскировки

Появление инструментов для "очеловечивания" ИИ-текста и постоянное совершенствование генеративных моделей делает их вывод всё более неотличимым от человеческого.

Фундаментальная проблема детекции

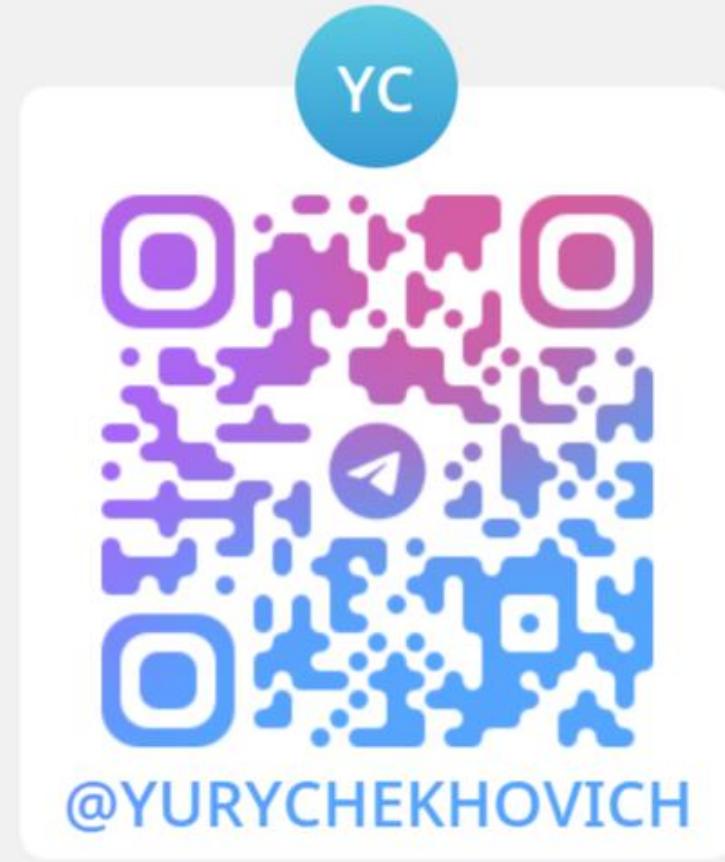
Принципиально не существует способа подтверждения корректности детекции ИИ-текста. Ситуация «слово алгоритма» против «слова автора»



Меры по снижению рисков

- **Обновление внутренней нормативной базы и авторских договоров:** запрет на использование произведения для обучения ИИ без согласия; отдельные положения для переводов, аудио, обложек
- **Политика ИИ «внутри дома»:** запрет редакторам/рецензентам загружать рукописи в публичные генеративные сервисы, требование **раскрывать AI-помощь** авторов (в колофонах/благодарностях или в метаданных)
- **Контроль качества вместо «детекторов»:** чек-листы фактчекинга для нон-фикшн, запрет на критические жанры (медицина, «опасные хобби») без профильного редактора/эксперта; не полагаться на AI-детекторы как на единственный инструмент
- **Антиспам-процессы** в приёме рукописей: ограниченные окна подачи, капчи/идентификация автора, быстрые «красные флаги» для типовых AI-слепков (уровень Clakesworld показал масштаб проблемы)
- **Защита бренда автора:** мониторинг маркетплейсов на «книги-двойники», готовые шаблоны претензий/DMCA, оперативная эскалация в KDP.
- **Веб-гигиена:** настроить robots.txt (блок GPTBot и др.) и заголовки X-Robots-Tag для AI; понимать ограничения этого подхода
- **Коммуникация с читателями и авторами:** объяснять, где ИИ помогает (редактура, обращение языка) и где нет (сюжет, выводы, факты), чтобы поддерживать доверие

Спасибо за внимание



Юрий Чехович
Институт проблем управления РАН
ведущий эксперт, к.ф.-м.н.

