

Генеративный искусственный интеллект и современный университет: НОВЫЕ ВЫЗОВЫ И НОВЫЕ ВОЗМОЖНОСТИ

**Т. А. Проворова,
Н. А. Сергеева,
Ю. В. Чехович,**

компания «Антиплагиат»,
г. Москва, Российская Федерация

Современное состояние генеративных сервисов

Генеративные сервисы на базе искусственного интеллекта стали неотъемлемой частью современной науки и образования. С одной стороны, они упростили процесс создания работ, дали авторам новые возможности, с другой – обострили проблемы оценки качества работ и степени самостоятельности авторов.

Отправная точка массового использования генеративных сервисов – конец ноября 2022 г., когда компания OpenAI объявила о запуске чат-бота ChatGPT [1], основанного на нейросетевой технологии InstructGPT. В течение очень короткого срока возникли десятки генеративных сервисов других компаний. Многие производители программного обеспечения объявили об аналогичных планах. Можно утверждать, что произошел технологический бум, объединенный общим понятием «искусственный интеллект».

Попробуем разобраться, почему современные генеративные сервисы обрели такую популярность, какие возможности они дают пользователям и какие проблемы создают для образования и науки. Рассмотрим подход к работе университетов в новых условиях, который позволит нивелировать проблемы и воспользоваться преимуществами новых технологий. Сосредоточимся на сервисах, которые генерируют именно текст, а сервисы, которые создают изображения, видео, звук и другие объекты, оставим для других исследований.

Причины популярности генеративных сервисов

Важно понимать, что разработки в области генеративных сервисов ведутся уже несколько десятков лет. Параллельно с этим развиваются и потребности в детекции искусственных текстов (первая публикация о детекторе искусственных текстов датируется 2008 г. [2]). Почему же последние разработки обрели такую популярность и стали использоваться массово?

Причин несколько. В первую очередь это низкий порог освоения результатов технологии, или, попросту говоря, доступность. Диалоговый режим работы, в котором пользователь может задавать вопросы на

естественном языке, не используя навыков программирования или специальных команд, создает ощущение, что общение идет с живым человеком.

Вторая причина – натуральность создаваемого текста. Результаты работы генеративных сервисов очень похожи на тексты, написанные человеком, и создается впечатление, что текст написан специалистом в предметной области заданного вопроса.

Еще одним достоинством, расширяющим сферу использования, оказалась возможность интерактивной работы с вопросами. Пользователь может уточнить свой запрос, дать новые вводные, попросить переписать ответ другими словами. Это позволяет корректировать и дорабатывать текст в соответствии с требованиями каждого конкретного случая.

Кроме того, генеративные сервисы существенно ускоряют процесс подготовки текста. Пользователь, обладая определенными навыками (которые вполне сопоставимы по сложности освоения с навыками использования поисковых интернет-систем), может в короткие сроки получать тексты значительного объема, что существенно экономит время и ускоряет процесс подготовки научных публикаций или написания учебных работ.

Вероятно, одним из самых существенных достоинств генеративных сервисов, с точки зрения студентов и ученых, оказывается формальная оригинальность создаваемых текстов. Разработчики генеративных моделей уделили особое внимание тому, чтобы генерируемые сервисами тексты не приводили пользователей к проблемам с авторскими правами.

Мы живем в эпоху, когда практически все работы на том или ином этапе подвергаются проверке на заимствования [3]. Поэтому текст, который по умолчанию не обнаруживается детекторами плагиата (такими как, например, система «Антиплагиат» [4]), особенно востребован.

Конечно, нужно отметить и недостатки генеративных сервисов. В первую очередь это вольное обращение с фактами. Сгенерированный текст зачастую содержит упоминания вымышленных событий, источников информации, организаций, документов и других объектов. Такого рода ошибки в текстах принято называть галлюцинациями. Их природа довольно понятна, если принимать во внимание основное назначение генерирующих моделей – создание текста. Разработчики моделей не уделяют внимания тому, чтобы генерируемый текст содержал правдивую информацию, им достаточно лишь правдоподобности результата.

Другой проблемой является неспособность современных моделей создавать относительно длинные связные тексты. Сервисы, которые отлично справляются с генерацией делового письма или новостной заметки, показывают довольно посредственные результаты, если запросить тексты значительного объема – научную статью, курсовую или дипломную работу, главу книги или диссертации. Такие тексты полны повторов, логически не связаны, не следуют какой-то общей мысли. Пользователи выходят из положения, разбивая работу с текстом на множество запросов и объединяя ответы. Однако это увеличивает трудоемкость и требует дополнительных усилий на поддержание связности результата.

Также можно охарактеризовать как недостаточный уровень работы сервисов с источниками информации. Пока сервисы не в состоянии самостоятельно подготовить научный обзор по узкой предметной области. Либо же результат такого обзора будет низкого качества.

В целом можно констатировать, что генеративные сервисы в их современной реализации позволяют быстро и относительно нетрудоемко готовить тексты, основное назначение которых – существовать. То есть тексты, которым другие люди не будут уделять серьезного внимания, вчитываясь в каждое слово и пытаясь уловить смысловые или эмоциональные оттенки. Нельзя не признать, что у таких текстов существует обширная область применения, в том числе в учебных и научных работах: статьях, дипломах, диссертациях, отчетах. Возможность быстро заполнить эти документы «водой» до нужного объема – наиболее очевидное объяснение популярности генеративных сервисов.

■ Этичное и неэтичное использование генеративных сервисов

Несмотря на то что популярную цель использования сервисов нельзя отнести к этичным, существует много корректных способов применения генеративных моделей в учебной и научной работе. Один из них – формирование структуры работы. Это помогает организовать свои мысли и идеи перед началом написания текста. При знакомстве с новой областью ИИ-чаты хорошо подходят для получения представления о терминологии. В этом случае сервисы традиционно работают очень неплохо, так как их модели обучены на огромном количестве примеров по самым разным темам.

Далее сервисы могут помочь при раскрытии текста по сформированной структуре. Автор здесь должен не просто получать и копировать текст в работу, а критически оценивать предложенное сервисом, т. е. применять собственный интеллект. В этом случае может быть создана работа высокого качества.

Генеративные сервисы активно используются для переводов и перефразировки, что помогает студентам работать с иноязычными материалами и расширять свой кругозор. Полезны возможности аннотирования источников и улучшения стилистики текста в тех случаях, когда автор готовит работу на иностранном языке.

Перечисленные способы позволяют этично использовать генеративный искусственный интеллект. Граница корректного применения проходит там, где заканчивается ответственность автора за создаваемую работу. Случаи, когда чат-боты используются, чтобы быстро «налить воды», а автор пытается быстро создать текст, который будет лишь похож на полноценное исследование, безусловно, относятся к неэтичному применению нейросетевых моделей. Аналогичным образом следует квалифицировать и использование возможностей сервиса по перефразировке текста для того, чтобы скрыть следы некорректных заимствований и затруднить работу детекторов плагиата.

Вообще можно ли называть плагиатом использование в своей работе сгенерированного текста? С одной стороны, тексты, сгенерированные нейросетью, как правило, плагиатом не являются, так как сами по себе не нарушают права других авторов. Плагиат подразумевает использование чужих идей или текстов без соответствующего разрешения, т. е. нарушение чужой интеллектуальной собственности. Если мы говорим об оригинальности, подразумевая, что текст не имеет текста-двойника в других работах, то текст, созданный искусственным интеллектом, является оригинальным. Однако в случае использования сгенерированного текста в научных или учебных работах возникает ряд нарушений, не относящихся к оригинальности, но касающихся этики и авторского права.

Самым существенным нарушением при подготовке работы является отсутствие указания на использование инструментов с алгоритмами искусственного интеллекта. Все больше вузов и научных журналов требуют декларировать все сервисы подобного рода, использованные при подготовке работы, а также цели и способы их применения. Есть примеры того, что вузы требуют прилагать к учебным работам скриншоты диалогов с чат-ботами, чтобы у преподавателей была возможность оценить корректность использования сервисов.

К менее очевидным нарушениям относится применение текста без соблюдения лицензионного соглашения с правообладателем сервиса. Мало кто читает текст соглашения при регистрации. Между тем некоторые сервисы прописывают в соглашении пункт о том, что права на тексты, произведенные этой нейросетью, принадлежат сервису. Используя этот текст без должной отсылки, пользователь нарушает права сервиса и лицензионное соглашение, с которым согласился.

Авторам, использующим искусственный интеллект, не стоит забывать, что именно они несут ответственность за все, что написано в их работе. И это не только выполнение этических стандартов, но и обеспечение качества и достоверности научных результатов.

■ Возможности обнаружения сгенерированного текста

Широкое распространение генеративных сервисов и относительно высокое качество результатов их работы делают актуальным вопрос: можно ли определить,

была ли учебная или научная работа написана автором самостоятельно или же ее создала нейросеть? Необходимо иметь инструмент, который поможет выявить фрагменты, качество текста которых необходимо оценить внимательно.

В системе «Антиплагиат» детектор машинно-сгенерированного текста появился в мае 2023 г. [5]. Предпосылкой к его созданию стали многолетние исследования специалистов в области детекции искусственных текстов, которые показали, что возможности детектирования существенно зависят от размера исследуемого документа [6]. Учебные и научные документы, как правило, содержат достаточное количество текста, чтобы алгоритмы детекции работали качественно.

Сервис детекции машинно-сгенерированного текста для принятия решения о том, кем написан текст, также использует нейросетевую модель, которую обучили специалисты компании. Для этого использовались десятки тысяч примеров естественных и искусственных текстов.

Система разбивает каждый проверяемый документ на фрагменты и классифицирует их, определяя, созданы они человеком или машиной. Решение принимается на основе сопоставления обширного набора признаков, позволяющих определить искусственное происхождение текста [7].

Как показано на рисунке, текст проверяемого документа разбивается на фрагменты, которые делятся на слова с помощью токенизатора, далее последовательности слов преобразуются в векторы большой размерности с помощью BERT-подобной языковой модели [8]. Затем векторы различных фрагментов анализируются совместно, уточняя результаты друг друга, и общая оценка поступает на классификатор, который принимает решение и представляет результат детекции. Пользователь в системе «Антиплагиат» видит специальный отчет, в котором содержатся отмеченные фрагменты, и может принять решение о дальнейшей судьбе работы.

■ Интеллект или нет?

После того как какой-то документ отмечен системой «Антиплагиат» как подозрительный с точки зрения наличия в нем сгенерированного текста, возникает вопрос: можно ли каким-то образом подтвердить или доказать, что текст действительно сгенерирован?

К сожалению, объективного ответа нет. Во всяком случае, пока. Сложность заключается в том, что не существует сколь угодно полных баз данных искусственных документов, вхождение текста в которые могло бы подтвердить искусственность. Человеческий экспертный анализ также не может дать полной гарантии. Как показывает практика, опытный преподаватель или редактор видит проблемные места текста, свойственные результатам генерации. К ним относятся фактические ошибки, низкая содержательная наполненность текста, повторы одних и тех же мыслей, логические противоречия, смена автором точки зрения и т. п. В то же время такие проблемы обнаруживаются и в написанных человеком текстах низкого качества. Поэтому мы рекомендуем экспертам в тех случаях, когда система обозначает документ подозрительным, не отклонять его только на основании этой отметки, а изучить текст и составить собственное мнение. Практически во всех ситуациях возникают содержательные претензии к документу, которые следует обсуждать с автором.

■ Работа университета в условиях генеративной реальности

Несмотря на оперативную реакцию систем обнаружения заимствований и обогащение их возможностями детекции искусственных текстов, университетам приходится соответствовать новым вызовам. Как же организовать научную работу в эпоху широкого использования генеративных сервисов?

Рассмотрим некоторые предложения по организации учебной и научной деятельности.

Во-первых, необходимо открыто признать возможность использования генеративных инструментов для подготовки работ. У студентов и исследователей должна быть возможность самостоятельно выбирать инструментарий при условии его корректного использования. При этом следует разработать и утвердить локальные акты вуза, которые определяют порядок применения инструментов в работе студентов и ученых. Важнейшей частью этого порядка является обязательная декларация полного перечня использованных инструментов с указанием частей работы, где они применялись, а также целей и способов использования. Нужно предусмотреть возможность включения в работу подтверждающих материалов.

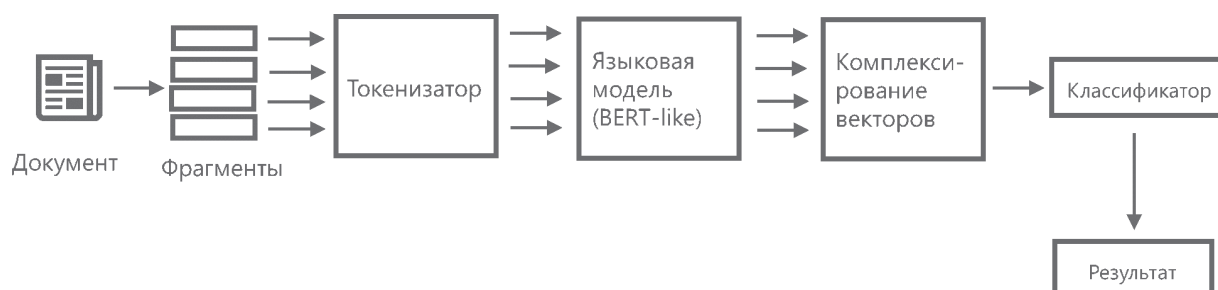


Рисунок. Схема работы алгоритма детекции машинно-сгенерированного текста

Во-вторых, важным аспектом организации научной работы является валидация результатов, полученных с помощью искусственного интеллекта. Для подтверждения их достоверности необходимо применять традиционные методы исследования. Это помогает предотвратить ошибочные выводы и обеспечить качество научных исследований.

В-третьих, возможность использования студентами сервисов на базе искусственного интеллекта не отменяет необходимость проверки работ на оригинальность. Современные детекторы плагиата не только проверяют работы на оригинальность, но и обнаруживают факты использования сгенерированного текста. Это можно сопоставлять с авторской декларацией применения такого текста.

Наконец, обучение и развитие навыков играют ключевую роль в успешной адаптации к использованию генеративных сервисов в научной работе. Включение в образовательные программы модулей по работе с искусственным интеллектом помогает развить критическое мышление и навыки оценки информации.

Стоит признать, что мы живем в новое время, и как было раньше, уже не будет никогда. Научить студентов правильно пользоваться сервисами генерации текста – значит дать им конкурентное преимущество на рынке труда. Не следует забывать и об обучении преподавателей работе с новыми инструментами. От их уровня погруженности будет существенно зависеть то, насколько корректно студенты будут пользоваться высокотехнологичными сервисами.

Немного о перспективах

Что будет дальше? Сегодня наука и образование стремительно развиваются. До недавнего времени мы даже не думали, что будем активно обсуждать влияние искусственного интеллекта на повседневную жизнь, и сложно прогнозировать, какие изменения произойдут в будущем. Однако можно предположить, что мы встретимся с различными сценариями, включая то, что созданный искусственным интеллектом текст будет неотличим от текста, написанного человеком, что заставит переосмыслить наши представления о реальности.

Возможно, будут введены общие нормативы по детектированию ИИ-текстов в научных работах и ис-

следователи будут обязаны соблюдать эти правила. Существует также перспектива введения требований к разработчикам сервисов, использующих чат-боты для генерации текстов, по обязательному применению системы логирования действий пользователей с возможностью последующего предъявления всей истории в случае необходимости.

Для сферы образования ключевым является своевременная реакция на изменения. Несмотря на то что это новый формат работы, прогресс требует его принятия. Важно, чтобы студенты знали, как вести себя в этих новых условиях. Задача преподавателей и администрации вузов – провести границы и описать их для студентов.

Список использованных источников

1. ChatGPT by OpenAI [Electronic resource]. – Mode of access: <https://chat.openai.com>.
2. Sameer Badaskar, Sachin Agarwal, and Shilpa Arora. 2008. Identifying Real or Fake Articles: Towards better Language Modeling. In Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II.
3. Чехович, Ю. В. Оценка корректности заимствований в текстах научных публикаций / Ю. В. Чехович, О. С. Бельская // Научное издание международного уровня – 2018: редакционная политика, открытый доступ, научные коммуникации: материалы 7-й междунар. науч.-практ. конф., Москва. – М.: ООО «Ваше цифровое издательство», 2018. – С. 158–162. – DOI: 10.24069/konf-24-27-04-2018.28. – EDN: XXXJQT.
4. Система распознавания интеллектуальных заимствований «Антиплагиат» / Ю. И. Журавлев [и др.] // Математические методы распознавания образов. – 2005. – Т. 12, № 1. – С. 329–332. – EDN: XUXXLТ.
5. Антиплагиат выявляет текст, созданный ChatGPT [Электронный ресурс] // Сайт компании «Антиплагиат». – Режим доступа: <https://antiplagiat.ru/news/text-chatgpt/>.
6. Gritsay, G. Automatic Detection of Machine Generated Texts: Need More Tokens / G. Gritsay, A. Grabovoy, Y. Chekhovich // 2022 Ivannikov Memorial Workshop (IVMEM), Moscow, Russian Federation. – 2022. – P. 20–26. – DOI: 10.1109/IVMEM57067.2022.9983964.
7. Поиск искусственно сгенерированных текстовых фрагментов в научных документах / Г. М. Грицай [и др.] // Доклады Российской академии наук. Математика, информатика, процессы управления. – 2023. – Т. 514, № 2. – С. 308–317. – DOI: 10.31857/S2686954323601677. – EDN GQRWLF.
8. A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT / Ce Zhou [et al.]. – 2023.

Аннотация

Бурное развитие генеративных сервисов, основанных на алгоритмах искусственного интеллекта, ставит серьезные вызовы перед системой высшего образования. Качество создаваемых ими искусственных текстов в значительной степени соответствует требованиям, которые предъявляются к учебным и научным работам. В статье анализируется этическое и неэтичное использование генеративных сервисов, описываются возможности детектирования искусственных текстов в системе «Антиплагиат», обсуждаются этические проблемы использования генеративных сервисов в научных и учебных целях. Авторы рассматривают варианты решения возникающих проблем.

Abstract

The rapid development of generative services built on artificial intelligence algorithms poses serious challenges to the higher education system. The quality of artificial texts corresponds to a great extent to the requirements of educational and scientific works. The article analyzes the ethical and unethical use of generative services, discusses the possibilities of detecting artificial texts in the “Antiplagiat” system, and considers ethical problems of using generative services for scientific and educational purposes. The authors consider options for solving the problems that arise.