2023;8(2):148–153 Научный редактор и издатель / Science Editor and Publisher

ИЗДАТЕЛЬСКАЯ ЭТИКА / PUBLISHING ETHICS

Дискуссионный документ. Перевод / Discussion document. Translation

https://doi.org/10.24069/SEP-23-22

Искусственный интеллект (ИИ) в принятии решений: дискуссионный документ*

Совет COPE

Версия 1. Сентябрь 2021 г.

COPE Discussion Document: Artificial intelligence (AI) in decision making

COPE Council

Version 1: September 2021

Дискуссионные документы COPE призваны привлекать внимание к актуальным (зачастую непростым) проблемам этики публикаций и давать импульс к обсуждению этих проблем. COPE приветствует комментарии, которые вносят вклад в текущую дискуссию.
Дискуссионные документы COPE не следует трактовать как выражение официальной политики COPE. Дискуссионные документы могут быть пересмотрены, но оставаться дискуссионными в контексте развития обсуждаемой проблемы и получения дальнейших комментариев от сообщества, или же они могут быть пересмотрены и стать официальной политикой COPE (после чего быть опубликованными как Руководство COPE).
Мы приветствуем дальнейшие комментарии: вы можете присылать свои отзывы исполнительному директору Natalie Ridgeway, cope_execofficer@publicationethics.org

Введение

За последние несколько лет технология искусственного интеллекта (ИИ) стремительно эволюционировала, и на ее основе в процесс публикации научных статей начали внедрять решения, основанные на данных. Инструменты ИИ уже используются в качестве вспомогательных на таких этапах публикационного процесса, как выбор журнала, определение темы статьи, определение соответствия темы статьи тематике журнала, рекомендации по привлечению рецензентов, оценка качества языка, выявление плагиата и дублирования текста, форматирование документов, а также оценка корректности планирования эксперимента и применения методов статистического анализа. Возможности использования ИИ в публикационном процессе быстро увеличиваются и влекут за собой новые проблемы и этические вопросы, которые необходимо учитывать. Данный документ представляет собой первую попытку ответить на эти вопросы и поделиться рекомендациями. В нем получает дальнейшее развитие первоначальная дискуссионная тема Форума COPE: «Искусственный интеллект (ИИ) в принятии решений» [1]. Цель данного дискуссионного документа – подготовить руководство и общие рекомендации по текущему состоянию вопроса.

Автоматизация versus ИИ

Существует четкое различие между автоматизацией и искусственным интеллектом, и эти два понятия не должны использоваться как взаимозаменяемые. Понимание того, что на самом деле представляет собой ИИ в публикационном процессе, является ключевым аспектом в этой дискуссии.

* Перевод статьи: COPE Council. COPE Discussion Document: Artificial intelligence (AI) in decision making. Version 1: September 2021. https://doi.org/10.24318/9kvAgrnJ

Под автоматизацией понимаются основанные на правилах компьютерные программы (*rules based software*), которые обеспечивают выполнение рабочего процесса на основе набора заранее установленных четких правил без вмешательства человека. Правила определяются разработчиками автоматизированной системы. Примером автоматизации процесса рецензирования является система программного обеспечения редакционного управления, которая автоматически отправляет по электронной почте напоминания авторам, задерживающим сроки повторной подачи рукописи, или система, удерживающая рукопись от обработки до следующего этапа рецензирования, если отсутствует какой-либо конкретный файл.

В случае ИИ речь идет о создании интеллектуальных систем, машин и программного обеспечения, которые могут имитировать интеллект и поведение человека. Цель ИИ – дополнить (и даже превзойти) человеческий интеллект, используя большие объемы данных для создания алгоритмов, нейронных сетей и графов, а также технологию глубокого обучения (*deep learning technology*) для достижения уровня «интеллекта», превосходящего по точности и масштабу человеческий разум. К направлениям ИИ мы относим обработку естественного языка (*natural language processing*, NLP) и машинное обучение (*machine learning*, ML). NLP – это технология, которая дает возможность компьютерам обрабатывать и пытаться понимать письменный или устный текст, а также выполнять такие задачи, как извлечение ключевых слов и классификация тем. ML – это направление, сосредоточенное на применении алгоритмов, обученных на наборах данных для выявления закономерностей, чтобы делать прогнозы, выполнять задачи и принимать решения, не прибегая к программированию.

Примером использования ИИ в рецензировании могут служить системы, которые предоставляют списки возможных рецензентов для рукописи, применяя технологию ML к большим базам данных исследователей и технологию NLP для обработки текста статей и отправки письма с просьбой принять участие в рецензировании рукописи наиболее подходящим кандидатам. Таким образом, ИИ дает рекомендации, которые мог бы дать редактор-человек, используя большую базу знаний и действуя в более глобальном масштабе.

Программное обеспечение на базе ИИ может предоставлять результаты, которые могут использоваться в автоматизированной системе или оцениваться и приниматься редактором или автором. Например, технология искусственного интеллекта может предсказать, в какой журнал с точки зрения тематики лучше подать рукопись. Затем система может сама, без участия человека, автоматически подать рукопись в этот журнал: это будет решением, основанным на ИИ. Другим примером являются результаты обнаружения перекрытия текста, предоставляемые таким программным обеспечением, как iThenticate [2], которые используются для принятия автоматического решения об отклонении рукописи в связи с выявлением плагиата без проверки человеком. Данные примеры показывают, каким образом автоматизация на основе ИИ может использоваться в принятии решений.

Таким образом, инструменты ИИ могут быть разработаны для предоставления людям рекомендаций на основе соответствующих данных; ИИ также может поддерживать полную автоматизацию некоторых процессов и принятие решений без вмешательства человека. В разделе «Этические дилеммы» мы рассмотрим этические последствия, связанные с принятием ИИ неконтролируемых решений в процессе публикации.

Зачем использовать автоматизацию и ИИ в издательском деле?

Автоматизация издательского процесса использовалась на протяжении десятилетий для обеспечения оперативного рецензирования рукописей без участия человека на каждом этапе этого процесса. К стандартным примерам автоматизации относятся системы, отправляющие напоминания авторам, рецензентам и редакторам о выполнении различных задач. Совсем недавно ИИ продемонстрировал свои возможности для решения задач, которые нелегко (а иногда и невозможно) решить человеческому разуму за приемлемый промежуток времени. Создаваемые инструменты ИИ и автоматизации могут повысить скорость и точность рецензирования. Программное обеспечение, созданное для обнаружения совпадения текста, благодаря перекрестной проверке миллионов документов обеспечивает уровень оценки, который не под силу человеческому мозгу. Возможности ИИ для распознавания образов (*pattern recognition*) позволяют обнаружить картели цитирования (*citation cartels*), манипуляции с изображениями [3], «нарезку салями» (*salami slicing*) и признаки «бумажной фабрики» (*papermill characteristics*) [4], а также все проблемные неэтичные практики в издательской деятельности, которые трудно идентифицировать.

2023;8(2):148–153 Научный редактор и издатель / Science Editor and Publisher

Новейшие технологии позволяют без участия человека оценить, соответствует ли рукопись стандартам качества в отношении языка, формата в соответствии с требованиями журнала, а также оформления рисунков и использования цитирования.

В публикационном процессе автоматизация используется, в основном, с целью обеспечить прохождение рукописей через рецензирование с минимальным участием человека в процессах, не связанных с принятием окончательного решения о публикации рукописи. Что касается ИИ, в настоящее время предпринимаются попытки предоставить ему право принимать окончательные решения о принятии или отклонении статей [5]. Мы полагаем, что такие инструменты ИИ могут помочь избежать личной предвзятости (*personal bias*), возникающей при вмешательстве редактора-человека (например, в случае, когда редактор предвзято относится к конкретным авторам или предпочитает привлекать рецензентов из определенных стран). Однако ИИ может иметь собственные проблемы, связанные с предвзятостью (*AI biases*), обусловленные данными, на которых ИИ обучался, действиями его разработчиков и самой конструкцией программного обеспечения [6; 7].

Общая цель использования автоматизации с помощью ИИ в публикационном процессе заключается в повышении качества автоматизации, снижении нагрузки на человека и ускорении процесса рецензирования, что в конечном итоге позволяет быстрее распространять проверенные и прошедшие рецензирование результаты исследований и снизить нагрузку на редакторов, рецензентов и авторов.

Контекст и проблемы

С развитием ИИ возникают этические вопросы, связанные с тем, когда и как его можно и нужно использовать для автономного принятия решений. Некоторые организации и правительства по всему миру предоставляют общие рекомендации по ответственному созданию и использованию ИИ [8–14]. Их ключевые положения можно отнести к трем основным группам:

– подотчетность (не дискриминирующая и справедливая): *accountability (non-discriminatory and fair)*;

– ответственность (участие и контроль человека): *responsibility (human agency and oversight*);

– прозрачность (техническая надежность и управление данными): *transparency (technical robustness and data governance*).

Непродуманная разработка и применение инструментов ИИ может привести к проблемам и нанести непреднамеренный вред. Как и в случае с любой другой технологией, для надлежащего и эффективного использования, широкого признания и извлечения значительной пользы необходимо ее понимать и тестировать. Поставщикам ИИ рекомендуется обеспечивать прозрачность и подотчетность своей деятельности, чтобы пользователи, в свою очередь, могли ответственно использовать ИИ-инструменты.

Одна из основных проблем подотчетности заключается в предвзятости. Алгоритмы ИИ обучаются на больших объемах данных, которые могут иметь присущие им предвзятости (*inherent biases*); также предвзятость может быть внесена в правила обучения, выбранные разработчиками. Важно знать о таких проблемах и быть готовыми либо исправлять их, либо обходить при разработке и применении инструментов ИИ. В тех случаях, когда это невозможно, в отношении ограничений и предвзятости крайне важно обеспечивать прозрачность. Сами алгоритмы должны регулярно проходить оценку эффективности модели для определения технической надежности и при необходимости переобучаться.

Доверие к ИИ является критически важным фактором, и развитие этих технологий должно включать в себя подход, ориентированный на людей, с учетом этического и культурного контекста пользователей ИИ, а также отдельных лиц, принимающих любые решения ИИ.

Этические дилеммы

Основные этические вопросы, которые поднимаются научным сообществом в связи с использованием ИИ для принятия решений в издательских системах, касаются трех ключевых аспектов: подотчетности, ответственности и прозрачности.

В этой связи возникают важные вопросы:

1. Существуют ли процессы, в которых допустима или даже желательна полная техническая автоматизация? Существуют ли процессы, в которых принятие решений с помощью ИИ было бы допустимым или желательным?

2. Существуют ли процессы, в которых полная автоматизация и/или принятие решений с помощью ИИ будут считаться неэтичными?

3. Какую информацию должны предоставлять журналы авторам (и рецензентам) об инструментах ИИ, используемых в их журнале? Насколько прозрачными должна быть издательская деятельность?

«Если машины будут участвовать в жизни человеческих сообществ в качестве автономных агентов, то от них будут ожидать соблюдения социальных и моральных норм сообщества.
Чтобы предоставить машинам такую возможность, необходимо определить эти нормы. Но чьи это нормы?»

Глобальная инициатива IEEE по этике автономных и интеллектуальных систем (*The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*), https://b.link/ethics

4. Что происходит, если автор или рецензент не согласен с решением или действием, принятым с помощью ИИ? Какие процедуры должны быть предусмотрены для обжалования действий ИИ?

5. Кто несет ответственность за решения, принятые с помощью ИИ? Как издатели могут обеспечить соблюдение и защиту прав человека?

Рекомендации

На данном этапе развития ИИ мы рекомендуем подходить к его внедрению с осторожностью. Ниже приведены рекомендации по этичному использованию ИИ при принятии решений в издательском процессе. Эти рекомендации адресованы издателям и редакторам, а также авторам научных статей.

Уточним, что в данном документе не рассматриваются вопросы этики создания, разработки и развития ИИ как такового. Данные вопросы обсуждаются в литературе многими ИТ-компаниями и более широкими структурами, рекомендуемую литературу можно найти в разделах ресурсов и списка литературы.

1. На данном этапе развития ИИ и изменения мировоззрений мы рекомендуем следующее: если речь идет о принятии окончательного решения в отношении статьи – принять к публикации или отклонить, – это решение должно приниматься с участием редактора. Такое решение не может быть принято исключительно ИИ.

2. Системы ИИ должны помогать людям делать информированный выбор в соответствии со своей ролью. Контроль со стороны человека – это ключ к обеспечению справедливости и соблюдению прав авторов при оценке их рукописей. В конечном итоге ответственность за редакционные решения, принятые как ИИ, так и редакторами-людьми, лежит на издателе.

3. Автоматизация с помощью ИИ для повышения скорости обработки, проверки, оценки качества и рецензирования может использоваться, считаться приемлемой и даже желательной во многих случаях, если в результате решение о принятии или отклонении рукописи не принимается исключительно ИИ. Например, в случае обнаружения ИИ-инструментом в рукописи рисунка с распознаваемым человеческим лицом без необходимой формы согласия, этот вопрос должен быть доведен до сведения редактора, который будет принимать решение об отклонении рукописи, либо автоматизированная система может сама отправить сообщение авторам с просьбой разъяснить ситуацию или предоставить соответствующую документацию.

4. Оценка неправомерного проведения исследований (*misconduct)* и добросовестности исследований (*research integrity*), приводящая к выражению озабоченности, ретракции или обращению в учреждения исследователей, также не должна основываться исключительно на решении ИИ.

5. Издателям следует обеспечить условия, в которых люди (редакторы, авторы и рецензенты), использующие технологию ИИ, доверяют ее результатам (как правило, на этапе тестирования) путем предоставления рекомендаций и технической поддержки, а также подробной информации о том, каким образом ИИ генерирует рекомендации. Издатели должны убедиться в том, что поставщики инструментов ИИ честны и откровенны в отношении того, как эти инструменты были созданы и обучены. Издателям следует сообщать обо всех обнаруженных предвзятостях в используемых алгоритмах или базах данных. Например, инструмент, предоставляющий данные о количестве цитирований автора, может быть полезен для оценки его показателей, но может предоставлять неточные данные из-за наличия самоцитирования. В таком случае либо необходимо информировать пользователей о данном ограничении, либо поставщик инструмента или издатель должны добавить параметр для подсчета самоцитирования.

6. Издателям следует оценить вероятность того, что используемые ими инструменты ИИ могут способствовать усилению и распространению предвзятости в отношении различных групп. Подаваемые материалы должны оцениваться с точки зрения содержания, а не расы, этнической принадлежности, пола, возраста, национальности или места проживания автора(ов). Инструменты, обученные на наборах исторических данных, должны при необходимости корректироваться на предмет искажений, связанных с предвзятостью. Издатели должны отслеживать подобные случаи и делиться всей релевантной информацией, что-

2023;8(2):148–153 Научный редактор и издатель / Science Editor and Publisher

бы на основе этой обратной связи разработчики могли обновлять свои инструменты.

7. Издателям следует принять меры для обеспечения прозрачности в отношении того, какие из их издательских процессов или этапов документооборота автоматизированы и в принятии каких решений задействован ИИ. Все участники процесса рецензирования – авторы, рецензенты и редакторы – должны быть четко информированы об использовании автоматизации с помощью ИИ, а также о том, как алгоритм пришел к тому или иному результату или выводу.

Авторам

1. Если у авторов имеются веские причины, они могут оспорить редакционное решение в соответствии со стандартными процедурами журнала. При оспаривании решения, которое было принято ИИ или основано на рекомендациях ИИ, следует придерживаться того же процесса, что и при решениях, принятых человеком. Независимо от того, было ли решение принято ИИ или редактором, журнал и издатель несут ответственность за принятое редакционное решение.

2. В случаях нарушений в принятии решений, несправедливого обращения или дискриминации авторам следует приводить веские аргументы при изложении своей позиции редакторам журнала или руководству издательства.

3. Авторы также имеют право на получение информации о том, какие издательские процессы или этапы документооборота в издательстве автоматизированы, и в принятии каких решений задействован ИИ.

Перспективы

В контексте развития ИИ и замены им некоторых традиционных систем возникает необходимость и четкое обязательство гарантировать, чтобы при его разработке принимались во внимание этические аспекты. По результатам глобального опроса об использовании ИИ и автоматизации, проведенного компанией *ARM Ltd*, доверие людей к решениям ИИ растет, особенно когда они видят доказательства того, что ИИ превосходит возможности человека. Прозрачность и подотчетность при разработке ИИ позволят людям доверять решениям ИИ и брать на себя ответственность за них. Основная проблема при использовании систем ИИ, которая долго оставалось без внимания и которую в настоящее время пытаются решить, связана с доступностью, качеством и консолидацией обучающих данных, а также с наличием предвзятостей в используемых наборах данных. Еще один важный вопрос заключается в том, как быстро необходимо переобучать инструменты ИИ на обновленных наборах данных, и как быстро они устаревают.

По мере развития технологий все больше инструментов ИИ будут наделены способностью к этическому рассуждению – одному из свойств человеческого поведения, которое они пытаются имитировать. Мы будем продолжать следить за ситуацией, а также за реакцией и дискуссией научного сообщества, чтобы регулярно обновлять рекомендации по этике использования ИИ в принятии решений в издательской деятельности.

Перевод: **Бюро переводов TextTranslate**

РЕСУРСЫ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА*

European approach to artificial intelligence – The EU's approach to artificial intelligence centres on excellence and trust, aiming to boost research and industrial capacity and ensure fundamental rights. Available at: https://b.link/digital-strategy

Everyday ethics for artificial intelligence – Digestible guide for developers to help reflect on ethical issues in their everyday job designing AI services. Available at: https://b.link/fundamentals

AI ethics guidelines global inventory – A list of frameworks that try to set out the principles of how systems for automated decision making can be developed and implemented ethically. Available at: https://b.link/algorithmwatch

Elaboration of a recommendation on the ethics of artificial intelligence – UNESCO artificial intelligence: towards a humanistic approach. Available at: http://b.link/ethics-ai

Ethics guidelines for trustworthy artificial intelligence, Chapter 3 – subsection of the guidelines containing an assessment list to help assess whether an AI system being developed, deployed, procured or used adheres to the seven requirements of trustworthy AI. Available at: http://b.link/altai

AI today, AI Tomorrow – Awareness, acceptance and anticipation of AI: A global consumer perspective. Available at: https://b.link/ai-survey

* Для Вашего удобства предоставляются ссылки на другие сайты, но Совет COPE не несет ответственности за содержание этих сайтов.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. COPE Forum 11 November 2019: Artificial intelligence (AI) in decision making. Available at: https://cope.onl/forum-ai
2. Crossref. similarity check. Available at: https://b.link/crossref
3. Bucci E. M. Automatic detection of image manipulations in the biomedical literature. *Cell Death & Disease*. 2018;9:400. https://doi.org/10.1038/s41419-018-0430-3
4. Cabanac G., Labbé C., Magazinov A. Tortured phrases: A dubious writing style emerging in science. Evidence of critical issues affecting established journals (preprint). Available at: https://b.link/tortured-phrases
5. Checco A., Bracciale L., Loreti P., Pinfield S., Bianchi G. AI-assisted peer review. *Humanities and Social Sciences Communications*. 2021;8:25. https://doi.org/10.1057/s41599-020-00703-8
6. Golden J. AI has a bias problem. This is how we can solve it. World economic Forum. Jan. 18, 2019. Available at: https://b.link/human-bias
7. Omowole A. research shows AI is often biased. Here's how to make algorithms work for all of us. World economic Forum. July 19, 2021. Available at: https://b.link/ai-machine
8. Institute of Electrical and Electronics Engineers. Ethics in action in autonomous and intelligent systems. Available at: https://b.link/ethics-action
9. IBM institute for Business Value. Advancing AI ethics beyond compliance. Available at: http://b.link/advancing-ai
10. House of Lords select Committee on Artificial intelligence, report of session 2017–19. AI in the UK: ready, willing and able? Available at: https://b.link/ai-committee
11. The Government of Canada. Responsible use of artificial intelligence (AI). Available at: http://b.link/responsible-ai
12. European Commission High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy AI. Available at: https://b.link/trustworthy-ai
13 UNESCO. UNESCO launches artificial intelligence needs assessment survey in Africa. Available at: https://www.unesco.org/en/articles/unesco-launches-artificial-intelligence-needs-assessment-survey-africa
14. Benjamins R. A choices framework for the responsible use of AI. *AI and Ethics*. 2021;1:49–53. https://doi.org/10.1007/s43681-020-00012-5

ВКЛАД АВТОРОВ

Разработка концепции:
Первоначальная дискуссия на форуме COPE по теме «Искусственный интеллект (ИИ) в принятии решений» была разработана в 2019 г. и проведена Heather Tierney (членом Совета COPE).
Итоговый дискуссионный документ был разработан и написан Marie Soulière (член Совета COPE) от имени Совета COPE.
Написание – подготовка исходного проекта (черновика): Marie Soulière
Написание – обзор и редактирование: Marie Soulière, Sonja Krane, Catriona Fennell, Howard Browman

БЛАГОДАРНОСТИ

Mattia Albergante, Nancy Chescheir, Sarah Elaine Eaton, Kim Eggleton, Suzanne Farley, Paul G Fisher, Alexander Laughery, Ana Marusic, Raymond Soulière, Siri Lunde Strømme и Laura Wilson рассмотрели и представили предложения по доработке документа.

Для цитирования: Совет COPE. Искусственный интеллект (ИИ) в принятии решений: дискуссионный документ. Версия 1. Сентябрь 2021 г. *Научный редактор и издатель*. 2023;8(2):148–153. https://doi.org/10.24069/SEP-23-22 (In Eng.: COPE Council. COPE Discussion Document: Artificial intelligence (AI) in decision making. https://doi.org/10.24318/9kvAgrnJ)

For citation: COPE Council. COPE Discussion Document: Artificial intelligence (AI) in decision making. Version 1: September 2021. https://doi.org/10.24318/9kvAgrnJ (Transl. in Russ.: *Science Editor and Publisher*. 2023;8(2):148–153. https://doi.org/10.24069/SEP-23-22).