УДК 004.032.26

Д. В. Коптев

студент кафедры прикладной информатики

Н. Н. Григорьева – старший преподаватель – научный руководитель

Анализ достоверности результатов работы поисковых систем

Искусственный интеллект (ИИ) – это область компьютерных наук, которая занимается созданием программ и систем, выполняющих задачи, требующие обычно человеческого интеллекта. Нейросети, способные генерировать изображения, аудио, тексты и более сложный контент, стали широко распространены среди обычных пользователей, благодаря развитию методов глубокого обучения и искусственного интеллекта в целом, и возникновению простых в использовании инструментов. [1]

1. Специфика генерации изображений

Можно выделить два основных способа генерации визуальных образов: [2]

1. Генерация изображения на основе текстового описания. Нейросеть получает на вход текстовое описание, учтя которое, получается готовое изображение.
2. Генерация изображения на основе существующего образа. Пользователь передаёт нейросети изображение и дальше ИИ меняет или дополняет его.

2. Общие проблемы и риски такого контента

Распространение ИИ, генерирующего контент, несёт в себе ряд проблем и рисков:

1. Нарушение конфиденциальности и безопасности. Происходит потеря и утечка в общий доступ огромного количества личной информации о пользователях, что несёт за собой размытие границ персональных данных и пренебрежительное отношение к ним.
2. Развитие киберпреступности. Рассматриваемая проблема ведёт к дистанцированию интернет-мошенника и жертвы, манипуляции общественным мнением, а также к автоматизации рутинных задач злоумышленника.
3. Отсутствие правовой базы. Перед государствами стоит проблема разграничения зон ответственности между разработчиком, оператором ИИ и самим ИИ.
4. Экономические проблемы. Главным риском в экономике является смещение спроса на рынке труда с кандидатов, занимающихся рутинным трудом, к кандидатам, выполняющим творческие, научные или прочие, трудно автоматизируемые, задачи.
5. Этические проблемы. Перед обществом встаёт проблема внедрения ИИ в обычную жизнь и необходимость изменения социальных паттернов взаимодействия людей на всё более частый контакт с машинами.

3. Подробное рассмотрение проблемы развития киберпреступности

Анализ достоверности особенно важен в аспекте этических проблем и проблем киберпреступности. Любая нейросеть должна быть обучена на некотором наборе данных. Обучающая выборка или готовый запрос составляется человеком в соответствии со своими вкусами и предпочтениями. Нейросети же действуют в соответствии с запросом, переданным изменяемым изображением, заданными настройками и обучающим массивом данных. Модели ИИ могут совершать случайно или заведомо неверные генерации. Итогом таких действий становятся ненастоящие изображения с ложной информацией, а также дипфейки (deepfake). Такой контент предоставляет богатый функционал для шантажа и махинаций, например в онлайн и оффлайн мошенничестве.

4. Распространение контента злоумышленника

Данный контент может распространяться почти всеми доступными средствами: передача распечатанного изображения или текста, расклейка и размещение подобного контента на улице, получение контента посредством социальных сетей, в переписке или в публикации пользователя, попадание на сгенерированное изображение или текст в контенте другого человека, обнаружение такого контента в поисковой выдаче браузера. Наибольший интерес вызывает рассмотрение проблемы выдачи недостоверного контента во время сёрфинга в интернете.

5. Работа поисковых движков

Работу поисковых систем можно обобщить до следующего процесса: Особые инструменты и алгоритмы производят поиск и сбор информации о документах, которые можно найти в интернете. Происходит анализ найденных страниц, нахождение гипертекстовых ссылок и занесение их в базу данных. [3] Пользователь вводит запрос в поисковую систему. В базу данных приходит запрос пользователя. По данному запросу выделяются соответствующие документы. Порядок выдачи документов определяется алгоритмами ранжирования, которые зависят от поисковой системы. Наиболее популярны алгоритмы ранжирования, которые размещают наиболее релевантную информацию выше в поисковой выдаче.

6. Сравнительный анализ наиболее популярных поисковых движков и их алгоритмов ранжирования

На какой позиции в картинках находится сгенерированное изображение при данных запросах.

Таблица 1 — Позиция сгенерированных изображений в картинках

	Pentagon explosion	Trump arrest	Selfie tank man china	Pope in jacket	Средний показатель (больше - лучше)

Google	1	3	1	1	1,50
Microsoft Bing	7	36 (0)	1	1	11,25 (3)
Yahoo!	44 (0)	5	1	1	12,75 (2,33)
DuckDuckGo	0	5	1	1	2,33
Baidu	5	1	0	6	4
Яндекс	1	1	6	1	2,25

Была построена диаграмма средних значений позиций сгенерированных изображений в разделе картинок. Результат представлен на рисунке 1.

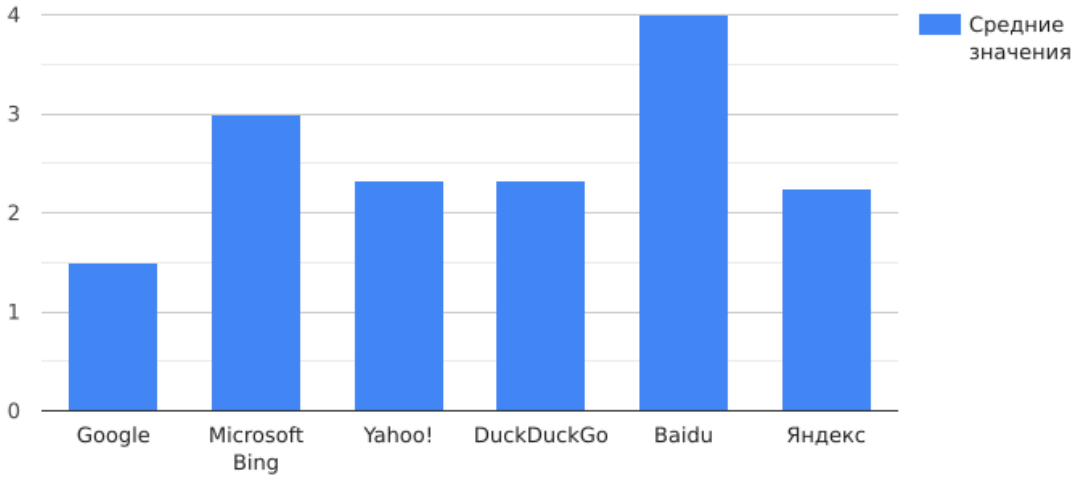

Рисунок 1 — Гистограмма средних значений

Были составлены таблицы количества статей, содержащих сгенерированные изображения, на первой странице выдачи разных поисковых систем по разным запросам. В таблице представлены оценки для следующих запросов: «Pentagon explosion», «Selfie tank man china», «Pope in jacket», «Trump arrest»

Таблица 2 — Количество страниц с ИИ по различным запросам

	ИИ на первой странице	Страниц с опровержениями	Картинок с маркировкой
Google	2/10, 4/10, 10/10, 0/10 Среднее = 5,33	10/10, 4/10, 10/10, 0/10 Среднее = 8	2/10, 3/10, 2/10, 0/10 Среднее = 2,33
Microsoft Bing	1/10, 3/10, 9/10, 0/10	3/10, 4/10, 9/10, 0/10	1/10, 1/10, 3/10, 0/10

	Среднее = 4,33	Среднее = 5,33	Среднее = 1,66
Yahoo!	1/10, 3/10, 9/10, 0/10 Среднее = 4,33	3/10, 4/10, 9/10, 0/10 Среднее = 5,33	1/10, 1/10, 3/10, 0/10 Среднее = 1,66
DuckDuckGo	2/10, 1/10, 10/10, 0/10 Среднее = 4,33	2/10, 1/10, 10/10, 0/10 Среднее = 4,33	1/10, 0/10, 3/10, 0/10 Среднее = 2
Baidu	3/10, 0/10, 1/10, 3/10 Среднее = 2,33	3/10, 0/10, 1/10, 3/10 Среднее = 2,33	1/10, 0/10, 0/10, 1/10 Среднее = 1
Яндекс	4/10, 4/10, 9/10, 1/10 Среднее = 4,5	10/10, 4/10, 10/10, 0/10 Среднее = 8	3/10, 4/10 1/10, 0/10 Среднее = 2, 66

7. Распространение контента в контексте поисковой выдачи

Рассмотрим наиболее частый сценарий попадания недостоверного контента в интернет. Сгенерированное изображение попадает на онлайн хостинги, а также вместе с недостоверными текстами на страницы интернет-статей. Поисковая система браузера находит сгенерированный ИИ контент по ключевым словам и фразам, не отличает его от реального и представляет пользователю в первых строчках выдачи. Эти ситуации вводят обычного пользователя в заблуждение.

8. Примеры

К сожалению, проблемы и риски распространения такого контента видны уже сейчас.

1. В мае 2023 года в социальных сетях начало стремительно распространяться фото, на котором запечатлено огромное чёрное облако над Пентагоном. Данное фото вызвало бурное обсуждение в сети и большой поток слухов и теорий. Минобороны США в кратчайшие сроки опровергло данное изображение, но ущерб не ограничился общественной паникой, вызванной в том числе распространением данных изображений верифицированными пользователями. Ряд крупных компаний понесли убытки на фондовом рынке, например такие фирмы как The Dow Jones Industrial Average и S&P 500. [4]
2. 24 марта 2023 года в соцсети Reddit было опубликовано реалистично сгенерированное изображение папы римского в богатом дизайнерском пуховике. [5] Картинка была опубликована в заведомо помеченном сообществе сгенерированных изображений. Но это не остановило подъём общественного резонанса и недовольства в связи с утечкой изображения и попадания его на иные ресурсы и интернет-страницы.
3. Похожая история произошла 2 апреля 2023 года в том же сообществе соцсети Reddit. На сайте появилось реалистичное изображение жертвы событий 1989 года на площади Тяньаньмэнь в Китае. [6] На данном изображение запечатлён человек снимающий селфи на

фоне едущего танка. Данное изображение было воспринято как насмешка, оскорбляющая чувства жертв данной трагедии.

9. Перспективы дальнейшего развития проблемы

Уже сейчас можно с уверенностью сказать, что будущее человечества будет связано с ИИ. В интернете уже содержится столько же сгенерированных изображений, сколько было снято фотографами со всего мира за 150 лет. [7] Производственные мощности растут, технологии улучшаются и появляются доступные решения, а значит таких изображений будет становиться больше с каждым годом. С развитием ИИ, алгоритмы генерации существенно преобразятся и, если сейчас опытный человек может отличить искусственное изображение от настоящего, то через пару лет распознать такой контент не сможет никто. Наиболее вероятна ситуация изменения интернет культуры и подхода к поиску информации. Скепсис по поводу происхождения того или иного контента заставит тщательнее выбирать источники информации и привыкать к сосуществованию с искусственными изображениями, аудио и текстами. Существует несколько решений данной проблемы, но, к сожалению, не существует ни одного без весомых издержек. В связи с этим можно сказать, что данная проблема не имеет правильного и единственного решения.

Список литературы

1. Петерс С.В. Нейросети для генерации изображений: области применения и юридические проблемы эксплуатации // Вестник науки. 2024. №3. С. 72.
2. Голышева Е.Н., Медведев А.А., Масалитин Н.С., Ильинская Е.В. Основные подходы к генерации изображений с помощью нейронных сетей // Инновационная наука. 2023. №11. С. 2.
3. Калмыков М. А., Медникова О. В. История развития поисковых систем и алгоритмы их работы // Научные известия. 2022. №28.
4. O'Sullivan D., Passantino J. CNN. URL: https://edition.cnn.com/2023/05/22/tech/twitter-fake-image-pentagon-explosion/ (дата обращения 11.04.2024)
5. Reddit. URL: https://www.reddit.com/r/midjourney/comments/120vhdc/the_pope_drip/ (дата обращения 11.04.2024)
6. Reddit. URL: https://www.reddit.com/r/midjourney/comments/129pcy8/selfie_tank_man_tiananmen_square_1989/ (дата обращения 11.04.2024)
7. Valyaeva A. Everypixel Journal. URL: https://journal.everypixel.com/ai-image-statistics (дата обращения 12.04.2024)