

XXII Всероссийской научно-практической конференции
КОРПОРАТИВНЫЕ БИБЛИОТЕЧНЫЕ СИСТЕМЫ:
ТЕХНОЛОГИИ И ИННОВАЦИИ

Модели генеративного искусственного интеллекта с полным их разоблачением

Юрий Викторович Чехович, к.ф.-м.н.

Исполнительный директор компании Антиплагиат

24–28 июня 2024 года
Санкт-Петербург, Великий Новгород

Бум генеративных сервисов

 YandexGPT

17 мая 2023 г.

 Writesonic
2020 г.

 Rytr

Июнь 2023 г.

 GIGA
CHAT

Апрель 2023 г.



30 ноября 2022

 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%	ПОСМОТРЕТЬ РЕЗУЛЬТАТЫ
<input type="checkbox"/> DOCX sample-2	16 Мая 2023 13:35	98,16%	ПОСМОТРЕТЬ РЕЗУЛЬТАТЫ

ПАПКИ Все документы

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

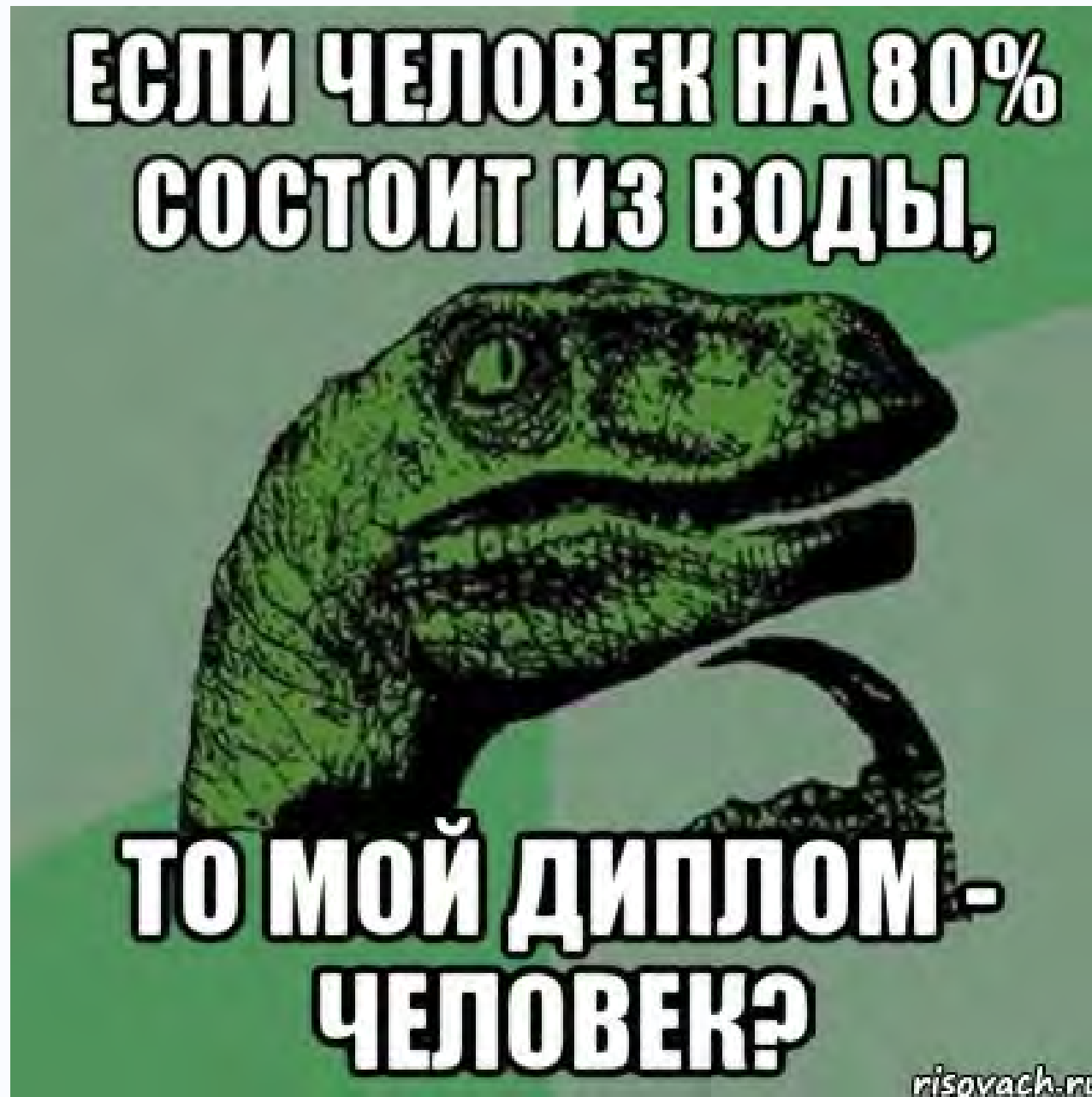
ПРОВЕРЕНО: 16.05.2023 10:42:10

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

Основной способ использования генеративного ИИ



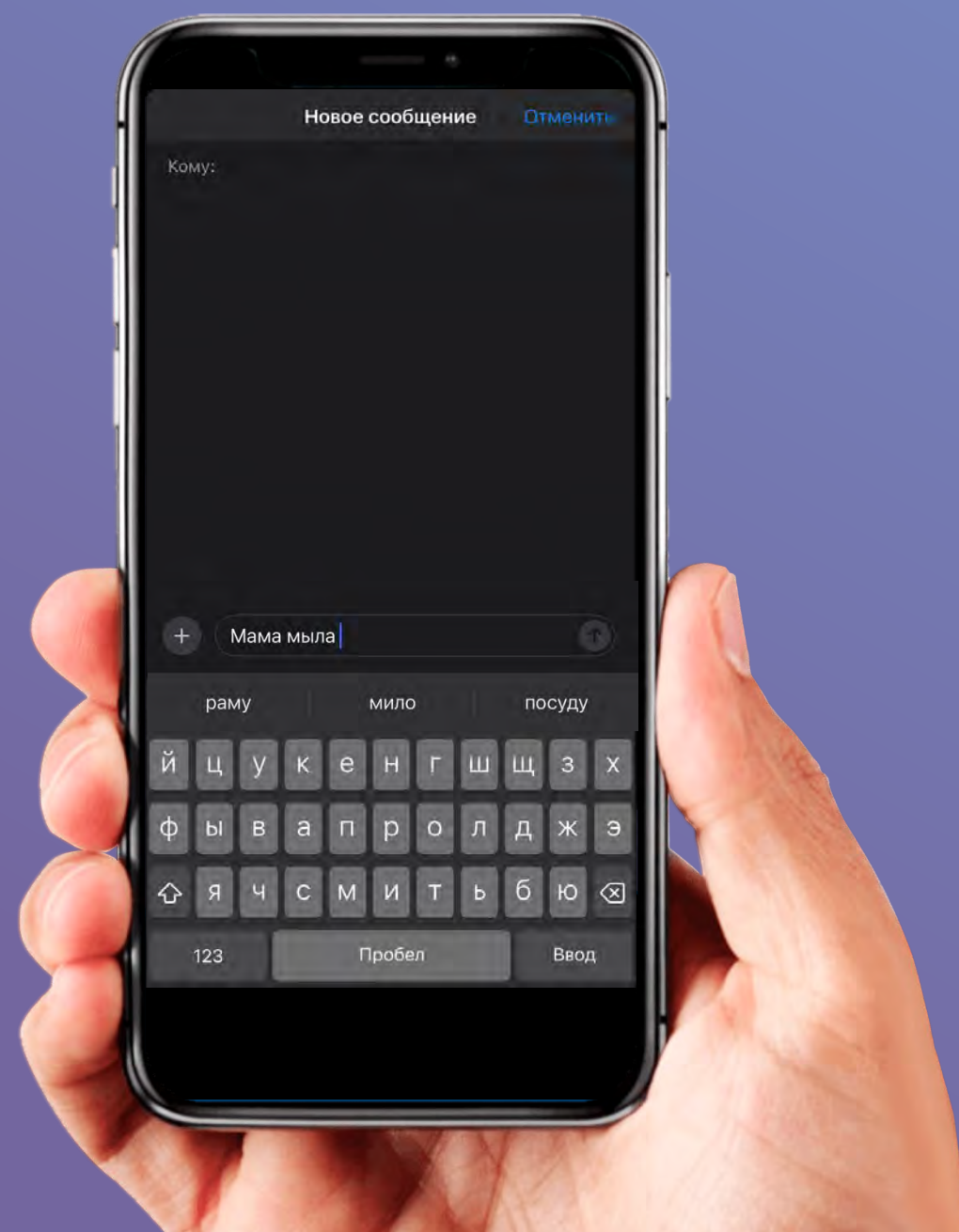
Почему генеративные сервисы востребованы?

- 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
- Чудесность – сложность моделей и кажущаяся недетерминированность результатов представляют сервисы как нечто магическое

Изучаем чудо: как работают генеративные сервисы?



T9



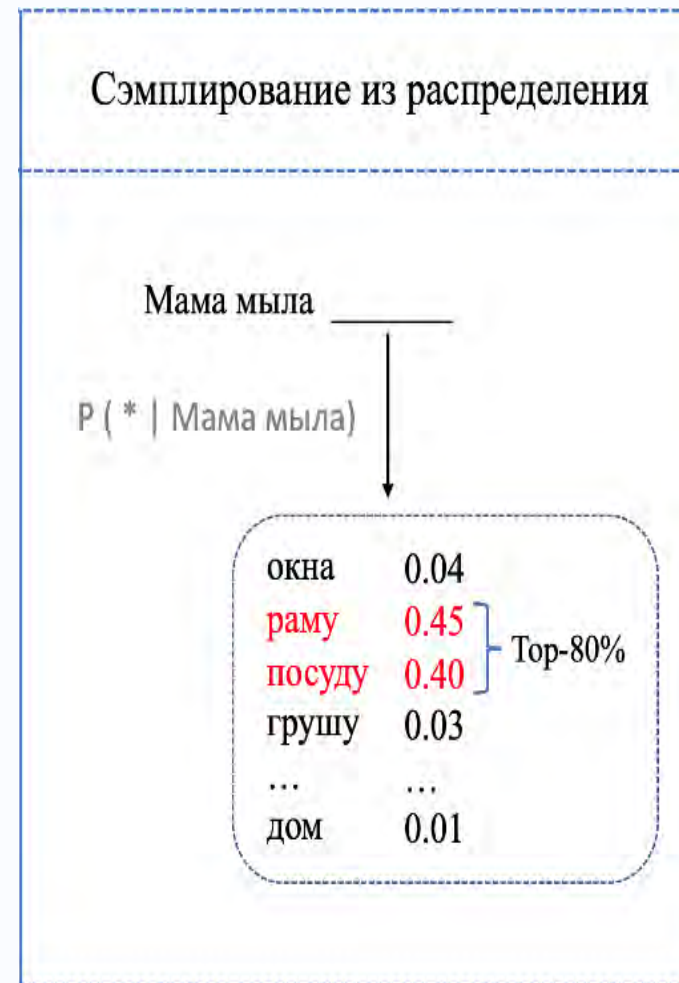
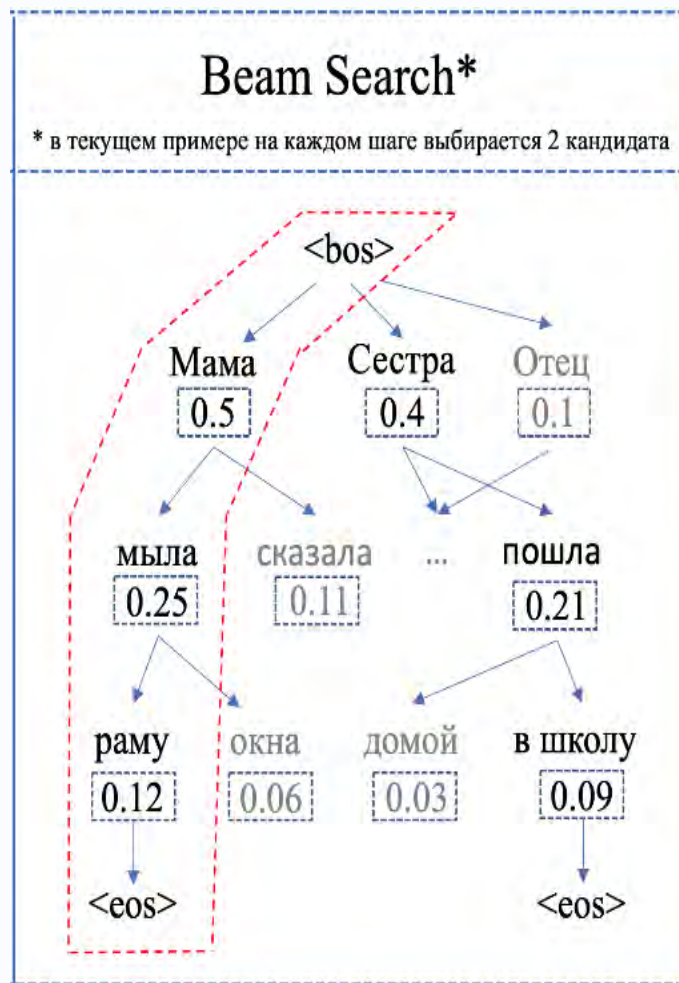
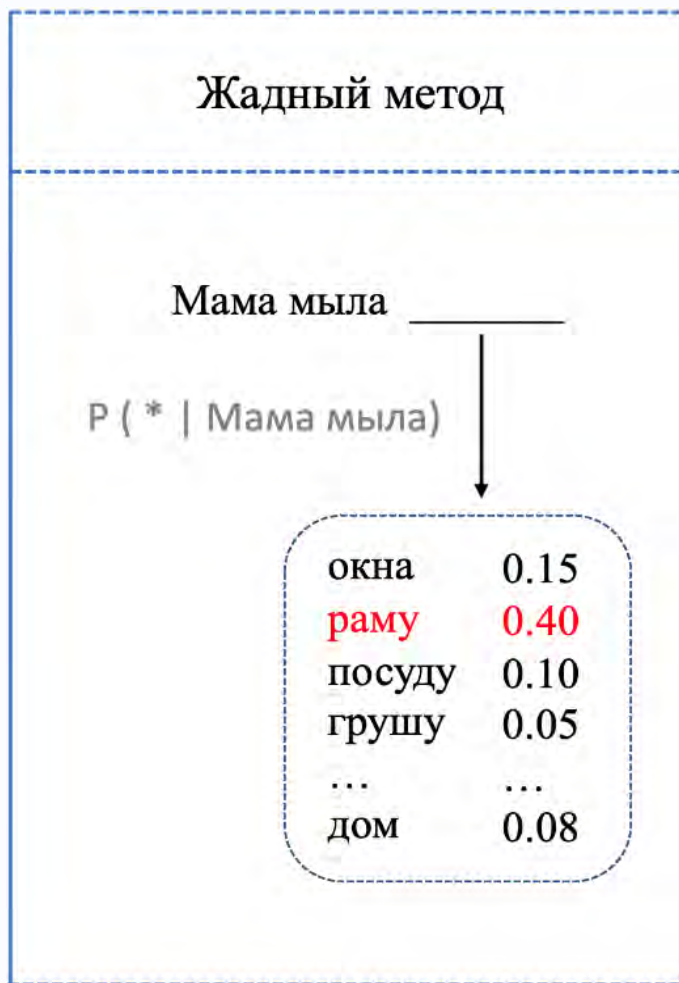
Как работают генеративные сервисы

Нужно проанализировать очень много текстов

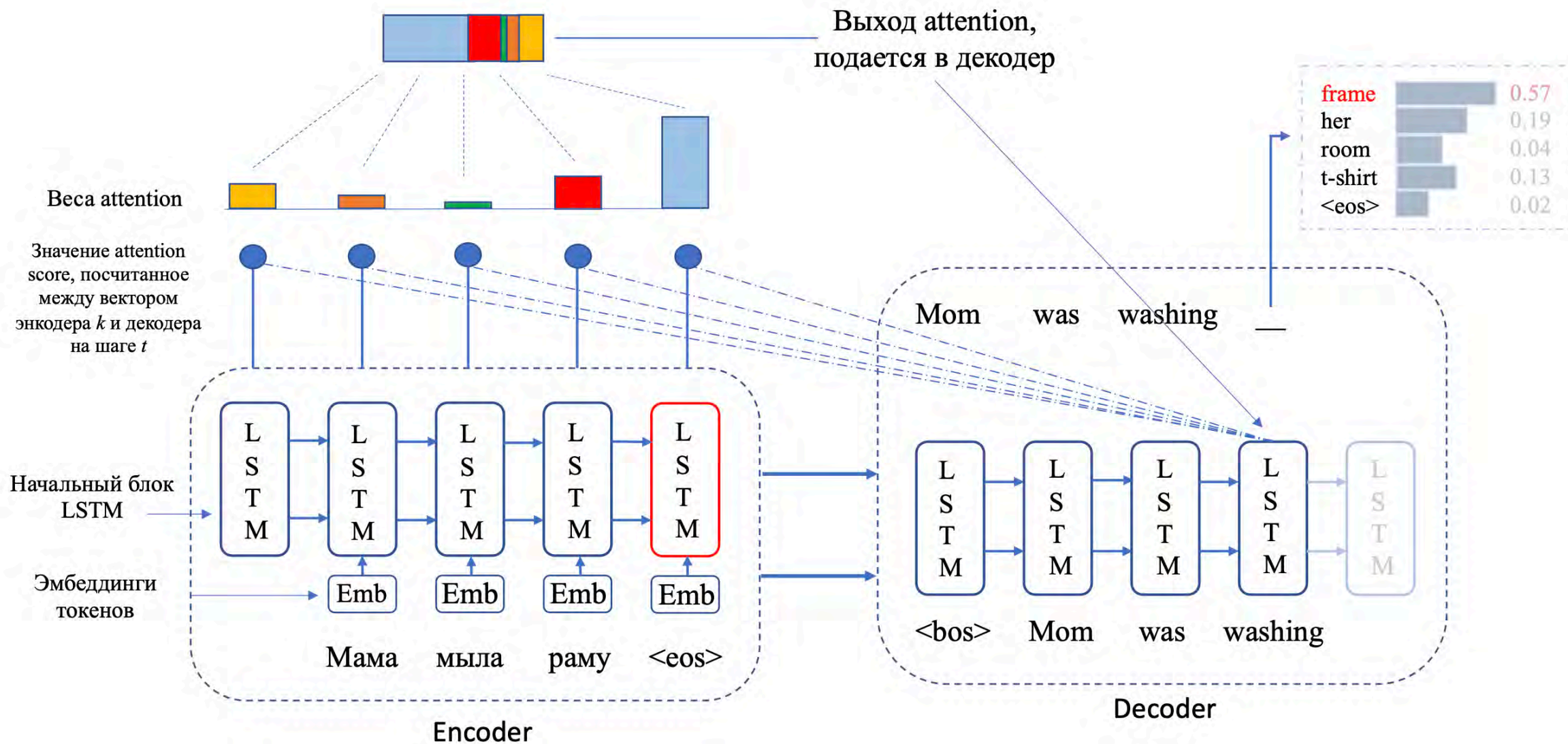


Генеративная модель – алгоритм продолжения текста

Мама мыла... [как продолжить?]



Генеративная модель – алгоритм продолжения текста



Суть 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.**

Тот факт, что современные генеративные модели ИИ решают некоторые задачи на «человеческом уровне», говорит скорее об уровне этих задач, чем о «силе» ИИ

Недостатки генеративных моделей

- «Галлюцинации» – недостоверные сведения

Ответ на вопрос «О чём книга "Муму?"»

“Муму” — это роман писателя Ивана Тургенева, опубликованный в 1954 году. Книга рассказывает историю простой девушки Герасима, которая была куплена богатой дворянкой и превращена в служанку. Главной душой и опорой Герасима в жизни становится его любимая корова Муму».

Недостатки генеративных моделей

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

Университетская книга



Ю. Чехович, А. Грабовой, Г. Грицай Модели генеративного искусственного интеллекта с полным их разоблачением // Университетская книга. – 2024. – № 5. – С. 58-65. – EDN YXSKBC.

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/text-chatgpt/>

ИИ или не ИИ

- Не существует объективных доказательств использования ИИ
- На что ориентироваться эксперту для подтверждения искусственности текста?

Признаки использования ИИ

- Несогласованность, нарушения связности

Как утверждает профессор Сямэньского университета Ло Сидон¹²⁵, основные международные политические теории на Западе, и реалистическая теория и либеральная теория ориентированы на национальный эгоизм, отражая то, что западные капиталистические страны считают абсолютную безопасность и интересы страны являются своими высшими целями.

растущая мощь других стран. Следовательно, это неизбежно будет способствовать формированию страны, которая с целью разрушения своего соседа. Если политика безопасности не будет поддерживать доверие в сфере

Китай стремится к поддержанию внутренней стабильности с помощью политических, экономических и социальных реформ. Это включает обеспечение социальной справедливости, справедливости и защиты прав и свобод граждан. Успехи в обеспечении стабильности внутри страны являются

Признаки использования ИИ

- Несогласованность, нарушения связности
- Много банальностей («эффект жвачки»)

Односторонние санкции, чрезмерное давление и война не решат проблему, а создадут новые. Китай активно использует свои культурные истоки, такие как конфуцианская философия и древние традиции, чтобы сформировать свою внешнюю политику. Он стремится к гармоничным и мирным отношениям с другими странами, основанным на принципах взаимного уважения, равноправия и сотрудничества.

Внутренняя политика Китая имеет огромное значение и непосредственную связь с международными отношениями. Успехи и направление внутренней политики Китая определяют его положение в мировом сообществе и влияют на его отношения с другими странами.

Признаки использования ИИ

- Несогласованность, нарушения связности
- Много банальностей («эффект жвачки»)
- Фразы, не несущие смысловой нагрузки

Тайваньский вопрос - типичный случай, когда на карту поставлены основные национальные интересы Китая. Независимо от того, какой страной она является, когда речь идет о двусторонних дипломатических отношениях с Китаем, она должна сначала принять принцип «одного Китая»²⁵⁷.

Многие страны прямо признают китайскую теорию «одного Китая» для внешнего мира. Любая страна должна четко признать принцип «одного Китая». Если принцип «одного Китая» не будет принят по вопросам основных национальных интересов, таким как вопрос Тайваня, двусторонние дипломатические отношения не могут быть сохранены.

интересов с конкретной политикой. Поэтому для анализа причин, по которым Китай в последние годы придает большое значение национальным основным интересам, важным возможным путем является сравнительный анализ границ между национальными основными интересами и не основными национальными интересами, а также четкое определение оперативной системы идентификации национальных основных интересов.

Признаки использования ИИ

- Несогласованность, нарушения связности
- Много банальностей («эффект жвачки»)
- Фразы, не несущих смысловой нагрузки
- Повторы («эффект рыбки Дори»)

национальных интересов. Существование и большая опасность антикитайских сил за рубежом, стоящих за вопросами «независимости Тайваня», «независимости Гонконга», «независимости Тибета» и «независимости Синьцзяна», а также вопросы прав человека — все это важные факторы основных интересов Тайваня. Существование и большая	важные факторы основных интересов Тайваня. Существование и большая опасность, исходящая от антикитайских сил за рубежом, стоящих за вопросами прав человека, являются ключевыми факторами включения вопросов Тайваня, Гонконга, Тибета и Синьцзяна в число основных интересов страны. По мере того, как Китай постепенно получает
с.68	с.68

128 Через ШОС, Китай сотрудничает с другими государствами-членами в пределах ШОС в различных сферах, включая политическую, экономическую, безопасности и культурную сферы. Китай активно предлагает и поддерживает идеи и проекты, которые способствуют сотрудничеству и укреплению безопасности в регионе. Одним из таких проектов является	129 Таким образом, Китай активно работает с другими странами через свою участие в ШОС и БРИКС, предлагая и поддерживая различные проекты и предоставляя существенные ресурсы, чтобы решить политические проблемы и способствовать экономическому развитию.
---	---

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

Китай также активно продвигает экологические программы, такие как создание национальных парков, восстановление экосистем и охрана биоразнообразия. Категорическая борьба с проблемами загрязнения

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
- Наличие ссылок в оригинальном тексте

эволюцию будущего мирового порядка¹⁹¹.

Противовесом в вопросе возвышения Китая является вопрос о нынешнем статусе США – являются ли они тем же гегемоном с сильнейшей позицией или происходят структурные изменения, снижающие их потенциал и позиции. Майкл Бекли, доцент Университета Тафтса, ставит под сомнение идею о том, что глобализация и ее продвижение привели к началу упадка США по сравнению с Китаем, и утверждает, что США не только не потеряли позиций по сравнению с Китаем, но и стали богаче, креативнее и мощнее в военном отношении со времен холодной войны¹⁹². Глобализирующийся мир и обязанности лидера не подорвали позицию Америки как великой державы, а скорее укрепили ее, потому что конкурентные преимущества Америки проистекают из ее гегемонистского положения, которые глобализация позволила использовать и получать максимальную выгоду от контроля международной системы¹⁹³.

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 определит веяние 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,

Можно ли использовать сгенерированный текст?

Рекомендации специалистов компании Антиплагиат:

- Не отклонять работы только на основании пометки о подозрительности
- Запрашивать дополнительные сведения при обнаружении искусственного текста (цели и способы использования, промты, диалоги с чатами, исходный текст до перевода)
- Обсуждать работу
- Совершенствовать нормативную базу организаций
- Стимулировать открытость промежуточных и финальных результатов

Спасибо за внимание!

Ваши вопросы

Юрий Чехович,

Исполнительный директор компании Антиплагиат, к.ф.-м.н.

chehovich@antiplagiat.ru

+7 495 223 23 84

8 800 777 81 28