Тема:

"Количественная оценка и минимизация галлюцинаций в больших языковых моделях (LLM)":

Название: Количественная оценка и минимизация галлюцинаций в больших языковых моделях (LLM)

Аннотация: В статье представлено исследование появления и сокращения галлюцинаций в больших языковых моделях. Вводятся новые количественные методы для оценки этих неточностей и обсуждаются различные стратегии их минимизации.

Ключевые слова: Большие языковые модели, Галлюцинации, Количественная оценка, Методы минимизации.

**1. Введение**

Значение Галлюцинаций в Больших Языковых Моделях (LLM)

Галлюцинации в больших языковых моделях (LLM), таких как GPT-3 и BERT, представляют собой случаи, когда модели генерируют информацию, не соответствующую реальности или фактам. Эти искажения могут проявляться в виде неверных утверждений, искаженной логики или нереалистичных выводов. Понимание и устранение галлюцинаций имеет ключевое значение для повышения надежности и точности LLM.

**Основные Аспекты и Влияние Галлюцинаций**

Влияние на Достоверность: Галлюцинации подрывают доверие к выводам LLM, что особенно критично в приложениях, требующих высокой точности, например, в медицине или юриспруденции.

Проблемы Безопасности: Галлюцинации могут вести к распространению недостоверной информации, что представляет угрозу для общественной безопасности и целостности данных.

Ограничения в Практическом Применении: Неверная информация, генерируемая моделями, ограничивает их использование в критически важных областях.

**Цели Исследований**

Определение и Классификация Галлюцинаций: Изучение различных типов галлюцинаций, их происхождения и механизмов появления.

Разработка Методов Количественной Оценки: Создание надежных метрик и инструментов для измерения степени и частоты галлюцинаций.

Методы Минимизации и Коррекции: Разработка стратегий для уменьшения частоты галлюцинаций, включая улучшение качества обучающих данных и алгоритмов обработки информации.

Повышение Общей Надежности LLM: Поиск способов улучшения моделей для предотвращения некорректного использования или интерпретации сгенерированных данных.

Изучение Влияния Галлюцинаций на Различные Приложения: Оценка воздействия галлюцинаций на различные сферы использования LLM, включая образование, бизнес и развлечения.

Это исследование направлено на улучшение понимания галлюцинаций в LLM и разработку эффективных методов их обнаружения и минимизации, что в конечном итоге повысит точность, надежность и полезность этих моделей в различных областях применения.

**2. Обзор литературы**

**Что такое модель GPT?**

GPT означает «Генераторный предварительно обученный преобразователь», первая обобщенная языковая модель в НЛП. Раньше языковые модели разрабатывались только для отдельных задач, таких как генерация текста, обобщение или классификация. GPT — это первая обобщенная языковая модель, когда-либо созданная в истории обработки естественного языка, которую можно использовать для различных задач НЛП. Теперь давайте рассмотрим три компонента GPT, а именно: генеративный, предварительно обученный и преобразователь, и поймем, что они означают.

Генеративные. Генеративные модели — это статистические модели, используемые для генерации новых данных. Эти модели могут изучать взаимосвязи между переменными в наборе данных для создания новых точек данных, аналогичных тем, которые есть в исходном наборе данных.

Предварительное обучение: эти модели были предварительно обучены с использованием большого набора данных, который можно использовать, когда обучение новой модели затруднено. Хотя предварительно обученная модель может быть не идеальной, она может сэкономить время и повысить производительность.

Трансформер: Модель трансформера, искусственная нейронная сеть, созданная в 2017 году, является самой известной моделью глубокого обучения, способной обрабатывать последовательные данные, такие как текст. Многие задачи, такие как машинный перевод и классификация текста, выполняются с использованием моделей преобразователей.

GPT может выполнять различные задачи НЛП с высокой точностью в зависимости от больших наборов данных, на которых он обучался, и его архитектуры из миллиардов параметров, что позволяет ему понимать логические связи внутри данных. Модели GPT, как и последняя версия GPT-3, были предварительно обучены с использованием текста из пяти больших наборов данных, включая Common Crawl и WebText2. Корпус содержит почти триллион слов, что позволяет GPT-3 выполнять задачи НЛП быстро и без каких-либо примеров данных.

**Обзор моделей GPT**

Модели GPT, сокращение от Generative Pretrained Transformers, представляют собой передовые модели глубокого обучения , предназначенные для генерации текста, похожего на человеческий. Эти модели, разработанные OpenAI, претерпели несколько итераций: GPT-1, GPT-2, GPT-3 и совсем недавно GPT-4.

Представленный в 2018 году GPT-1 был первым в этой серии, в котором использовалась уникальная архитектура Transformer, позволяющая значительно улучшить возможности генерации языков. Он был построен с использованием 117 миллионов параметров и обучен на наборах данных из Common Crawl и BookCorpus. GPT-1 может генерировать свободный и связный язык в некотором контексте. Однако у него были ограничения, в том числе склонность к повторению текста и трудности со сложным диалогом и долгосрочными зависимостями.

Затем OpenAI выпустила GPT-2 в 2019 году. Эта модель была намного больше, с 1,5 миллиардами параметров и обучалась на еще большем и разнообразном наборе данных. Его главной силой была способность генерировать реалистичные текстовые последовательности и человеческие ответы. Однако GPT-2 с трудом справлялся с сохранением контекста и связности в более длинных отрывках.

Внедрение GPT-3 в 2020 году стало огромным шагом вперед. Обладая ошеломляющими 175 миллиардами параметров, GPT-3 был обучен на обширных наборах данных и мог генерировать тонкие ответы при выполнении различных задач. Он может генерировать текст, писать код, создавать изображения и многое другое, что делает его ценным инструментом для многих приложений, таких как чат-боты и языковой перевод. Однако GPT-3 не была идеальной и имела свою долю предубеждений и неточностей.

Вслед за GPT-3 OpenAI представила обновленную версию GPT-3.5 и в конечном итоге выпустила GPT-4 в марте 2023 года. GPT-4 — это новейшая и наиболее совершенная из многомодальных языковых моделей OpenAI. Он может генерировать более точные утверждения и обрабатывать изображения в качестве входных данных, позволяя создавать подписи, классификации и анализ. GPT-4 также демонстрирует творческие способности, такие как сочинение песен или написание сценариев. Он поставляется в двух вариантах, различающихся размером контекстного окна: gpt-4-8K и gpt-4-32K.

Способность GPT-4 понимать сложные подсказки и демонстрировать человеческое поведение при выполнении различных задач — это значительный шаг вперед. Тем не менее, как и в случае со всеми мощными инструментами, существуют обоснованные опасения по поводу потенциального неправильного использования и этических последствий. Крайне важно учитывать эти факторы при изучении возможностей и приложений моделей GPT.

Предыдущие исследования галлюцинаций в LLM.

Существующие методы количественной оценки.

**3. Методология**

Обзор Литературы: Применение Больших Языковых Моделей (LLM) в Финансах и Проблемы с Галлюцинациями

Использование больших языковых моделей (LLM) в финансовой сфере обретает всё большую популярность благодаря их способности к анализу больших объемов данных, прогнозированию трендов и автоматизации задач. Однако проблемы, связанные с галлюцинациями, оказывают значительное влияние на надежность и точность этих моделей в финансовом контексте.

**Применение LLM в Финансах**

Автоматизация и Помощь в Принятии Решений: LLM используются для анализа финансовых отчетов, новостей и прогнозирования рыночных тенденций.

Обработка Естественного Языка для Финансового Анализа: Модели обрабатывают большие объемы текстовой информации, позволяя выявлять закономерности и получать ценные инсайты из новостей и финансовых отчетов.

**Проблемы Галлюцинаций в Финансовом Контексте**

Неточные Прогнозы и Анализ: Галлюцинации могут привести к неправильным выводам и анализу данных, что критично в финансах.

Риски для Инвестиций: Неверная интерпретация данных может ввести в заблуждение инвесторов, повлиять на рыночные тенденции и финансовые решения.

**Акцент на Литературе, Исследующей Эти Проблемы**

Исследования Эффектов Галлюцинаций в Финансах: Анализ научных работ, посвященных влиянию неточностей и галлюцинаций LLM на финансовый анализ и прогнозирование.

Стратегии Преодоления Галлюцинаций: Обзор литературы о методах минимизации влияния галлюцинаций в финансовых приложениях LLM, включая улучшенные алгоритмы обработки данных и методы верификации информации.

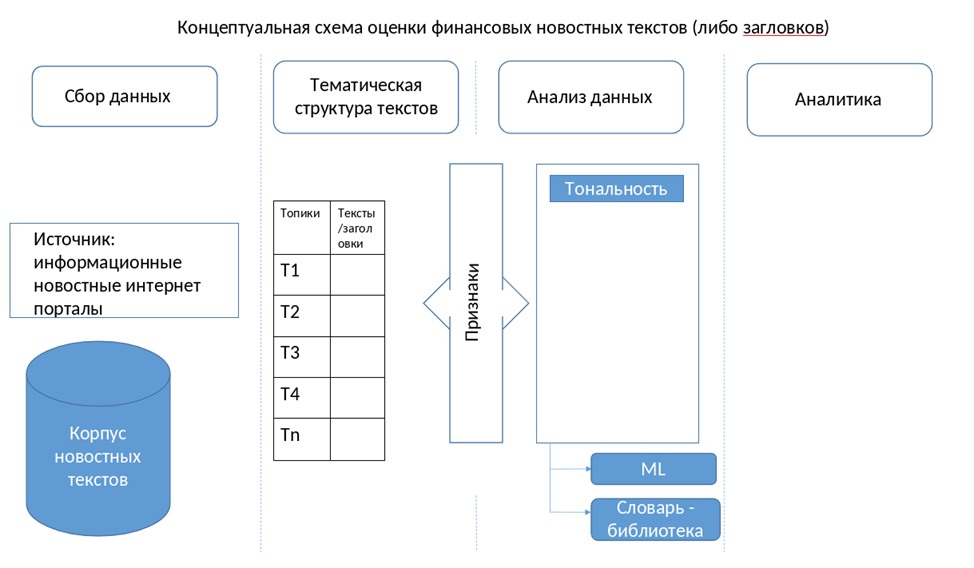
**Выводы и Направления Дальнейших Исследований**

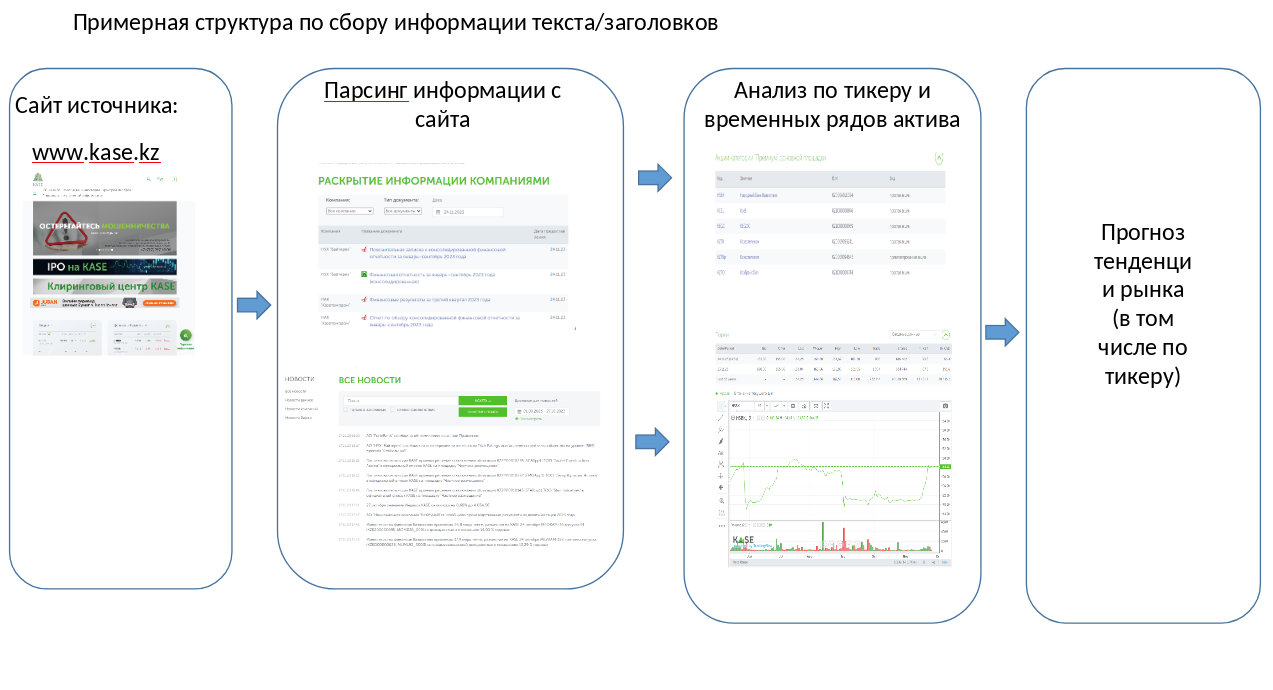
Обзор литературы подчеркивает значимость точности и надежности LLM в финансовом секторе. Исследования в этой области должны сосредоточиться на разработке решений для минимизации риска галлюцинаций, что повысит ценность и безопасность применения LLM в финансах.

**4. Результаты**

Порядок создания прототипа Большой языковой модели ( БЯМ):

1. Описание структурной схемы поиска
2. Создание базовой модели LLM, например, GPT-1.
3. Производится проверка на массиве информации, например BBC news
4. Проводится расчет совпадений токенов





**Заключение**

**Список использованной литературы:**

|  |  |  |
| --- | --- | --- |
| № | Наименоваение темы статьи | веб линки |
| 1 | Forecasting stock prices using LLMs with financial news | <https://community.wolfram.com/groups/-/m/t/2959055> |
| 2 | Large Language Models in Finance: A Survey | <https://www.researchgate.net/profile/Yinheng-Li> |
| 3 | Sentiment analysis of stock market news with LLM | <https://medium.com/@sashkoi1234/sentiment-analysis-of-stock-market-news-with-llm-part-1-1a6055b769f4> |
| 4 | Применение искусственного интеллекта на финансовом рынке. | <https://www.cbr.ru/Content/Document/File/156061/Consultation_Paper_03112023.pdf> |
| 5 | BloombergGPT: A Large Language Model for Finance | <https://arxiv.org/pdf/2303.17564.pdf> |
| 6 | FinGPT: Open-Source Financial Large Language Models | <https://tidwellgroup.com/navigating-the-future-of-finance-how-large-language-models-are-reshaping-the-industry/> |
| 7 | AI in Quantitative Finance: A New Dawn with Large Language Models | <https://medium.com/@yousra.aoudi/ai-in-quantitative-finance-a-new-dawn-with-large-language-models-290e3da32344> |
| 8 | Integrating Stock Features and Global Information via Large Language Models for Enhanced Stock Return Prediction | <https://arxiv.org/abs/2310.05627> |
| 9 | Slope TransFormer: The first LLM trained to understand the language of banks | <https://medium.com/slope-stories/slope-transformer-the-first-llm-trained-to-understand-the-language-of-banks-88adbb6c8da9> |
| 10 | FinChat | <https://preview.mailerlite.io/emails/webview/420148/105382283927619173> |
| 11 | Introducing BloombergGPT, Bloomberg’s 50-billion parameter large language model, purpose-built from scratch for finance | <https://www.bloomberg.com/company/press/bloomberggpt-50-billion-parameter-llm-tuned-finance/> |
| 12 | Integrating Stock Features and Global Information via Large Language Models for Enhanced Stock Return Prediction | <https://arxiv.org/pdf/2310.05627.pdf> |
| 13 | linkedin | <https://www.linkedin.com/pulse/finllm-unleashed-exploring-potential-financial-large-language/> |
| 14 | FinGPT: Democratizing Internet-scale Data for Financial Large Language Models | <https://arxiv.org/abs/2307.10485> |
| 15 | Slope TransFormer: The first LLM trained to understand the language of banks | <https://medium.com/slope-stories/slope-transformer-the-first-llm-trained-to-understand-the-language-of-banks-88adbb6c8da9> |