Estado del Arte de LLM y Predicción Bursátil con Sistema Multiagente

Elaborado por:

Salvador Nicolás Sanchez

Índice

Estado del arte de los LLM

El análisis del estado del arte se centra en explorar las aplicaciones actuales de agentes inteligentes que hacen uso de los 'Large Language Models' (LLM). En esta investigación, se llevó a cabo una exhaustiva revisión de diversos papers y artículos relacionados con la actualidad de los modelos grandes de lenguaje, con el objetivo de contextualizar y comprender a fondo las aplicaciones prácticas que estos modelos ofrecen. A través de esta revisión bibliográfica, se busca proporcionar una visión integral de las características de los LLM y cómo se aplican en la práctica, destacando las innovaciones y tendencias emergentes en este campo.

"A Survey on Evaluation of Large Language Models"

YUPENG CHANG (estudiante de la Escuela de Inteligencia Artificial, Universidad de Jilin, China), 17/10/2023.

Este paper nos permitió comprender algunas de las aplicaciones prácticas que tienen los LLM, como la mejora del procesamiento de lenguaje natural, el diagnóstico médico, la resolución de problemas matemáticos, la recomendación de contenido y otras áreas emergentes que aprovechan la capacidad de estos modelos para entender y generar texto de manera avanzada.

A su vez, también nos compartió los aspectos éticos que se enfrentan a la hora de generar un nuevo LLM. Se exploraron desafíos relacionados con sesgos inherentes en los datos de entrenamiento, la equidad en la aplicación de estos modelos y las consideraciones éticas en su implementación en diversos campos, subrayando la importancia de abordar estas cuestiones para garantizar un desarrollo y uso ético de los LLMs.

Por último, el papper también abordó las tendencias actuales en este ámbito. Entre ellas se destacan la capacidad de generalización a dominios no vistos, la interacción multimodal que combina texto con otros tipos de datos, y la aplicación cada vez más frecuente de LLMs en entornos educativos, sugiriendo que estos modelos están evolucionando y siendo adoptados en una variedad de contextos de manera continua. Además, se exploraron estrategias innovadoras en el diseño de prompts y se discutieron las posibles direcciones futuras de investigación y desarrollo en el campo de los LLMs.

"The imperative for regulatory oversight of large language models (or generative AI) in healthcare"

Bertalan Mesko y Eric J. Topol, 06/07/2023.

Este artículo examina detalladamente los desafíos y riesgos asociados al uso de Large Language Models (LLMs) en el ámbito de la atención médica, centrándose en tecnologías como el ChatGPT de OpenAI. Aunque estos modelos de lenguaje avanzados ofrecen potencial en medicina, surgen preocupaciones regulatorias, éticas y prácticas que requieren atención.

El texto aborda varios desafíos clave, como la protección de la privacidad de los datos de pacientes, la determinación de responsabilidades en casos de daño a pacientes debido a recomendaciones de IA, y la gestión de la propiedad intelectual cuando los LLMs generan contenido similar a investigaciones médicas propietarias. Se destaca la necesidad de regulaciones para garantizar la calidad y consistencia de los consejos médicos generados por la IA, así como la importancia del consentimiento informado de los pacientes al utilizar herramientas de IA en su atención médica.

La transparencia y la interpretabilidad de los modelos de IA, la equidad y la prevención de sesgos, la propiedad de los datos de entrenamiento y el riesgo de dependencia excesiva en la IA también son áreas críticas discutidas en el artículo. Se plantea la posibilidad de establecer una nueva categoría regulatoria para LLMs, junto con la necesidad de orientación reglamentaria para empresas y organizaciones de atención médica. En resumen, se enfatiza la importancia de una implementación ética y responsable de los LLMs para preservar la confianza de pacientes y profesionales de la salud.

"Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection"

Kai Greshake (alumno de la universidad de Saarland), 30/11/2023

Este informe se centra en aspectos cruciales de la seguridad en las aplicaciones que incorporan modelos de lenguaje grande (LLM). El enfoque principal de la investigación se dirige hacia un tipo específico de amenaza denominada "Indirect Prompt Injection". Se exploran casos concretos de compromisos, incluyendo la vulneración de usuarios y motores de búsqueda, lo que destaca la relevancia y gravedad de esta amenaza en escenarios del mundo real.

Proporciona una contextualización técnica profunda al referirse a conceptos clave, como "LLM-Integrated Applications" y "Search Engine", lo que sugiere que la investigación se sumerge en el análisis detallado de la interacción entre estos elementos. La metodología utilizada para llevar a cabo ataques de inyección indirecta de prompts se describe con el fin de comprender mejor cómo los adversarios pueden aprovechar vulnerabilidades en las aplicaciones integradas con modelos de lenguaje.

Asimismo, la inclusión de referencias extensas en el informe demuestra un enfoque integral en la revisión y el análisis de la seguridad de modelos de lenguaje, abarcando temas que van desde la ética y la transparencia hasta la aplicación práctica de tecnologías emergentes. La lista de referencias proporciona una base sólida de conocimientos previos y trabajos relacionados en el campo de la seguridad de lenguaje natural y modelos de lenguaje.

Problemática

En el estudio de los mercados existe un constante objetivo de lograr predecir cómo van a fluctuar los precios en la bolsa, esto ha llevado a muchos análisis de la economía con un enfoque técnico con modelos que en base a los valores anteriores de las acciones, bonos o cotizaciones buscan predecir su futuro precion, sin embargo este ámbito es muy sensible a los rumores, noticias y hechos de las personas que lo rodean, lo que hace importante un modelo que logre aportar al análisis del mercado un análisis de lo humano. En este ámbito planteamos un modelo que en base a lo que se hable en la sociedad logre predecir su impacto sobre la economía.

Planteo

Este modelo se basa en que una noticia, rumor o dicho va a generar una reacción en los remitentes que se refleja en la confianza de los bonos y por consiguiente en su valor. También suponemos que disponemos de la información de las noticias de diferentes portales, chat de diferentes grupos públicos en diferentes redes sociales y de diferentes ámbitos u otro tipo de canal donde se comparta información hacia las personas como lo pueden ser la televisión, canales de youtube y demás.

La capacidad para analizar datos provenientes de estas múltiples fuentes nos permite obtener una perspectiva más completa y contextualizada del entorno social, permitiendo evaluar no solo la magnitud de la información, sino también su impacto potencial en la confianza del mercado. En última instancia, esta aproximación integral busca mejorar la precisión de las predicciones financieras al considerar no sólo los eventos en sí, sino también la complejidad de las interacciones y percepciones que tiene la sociedad actual.

Estructura

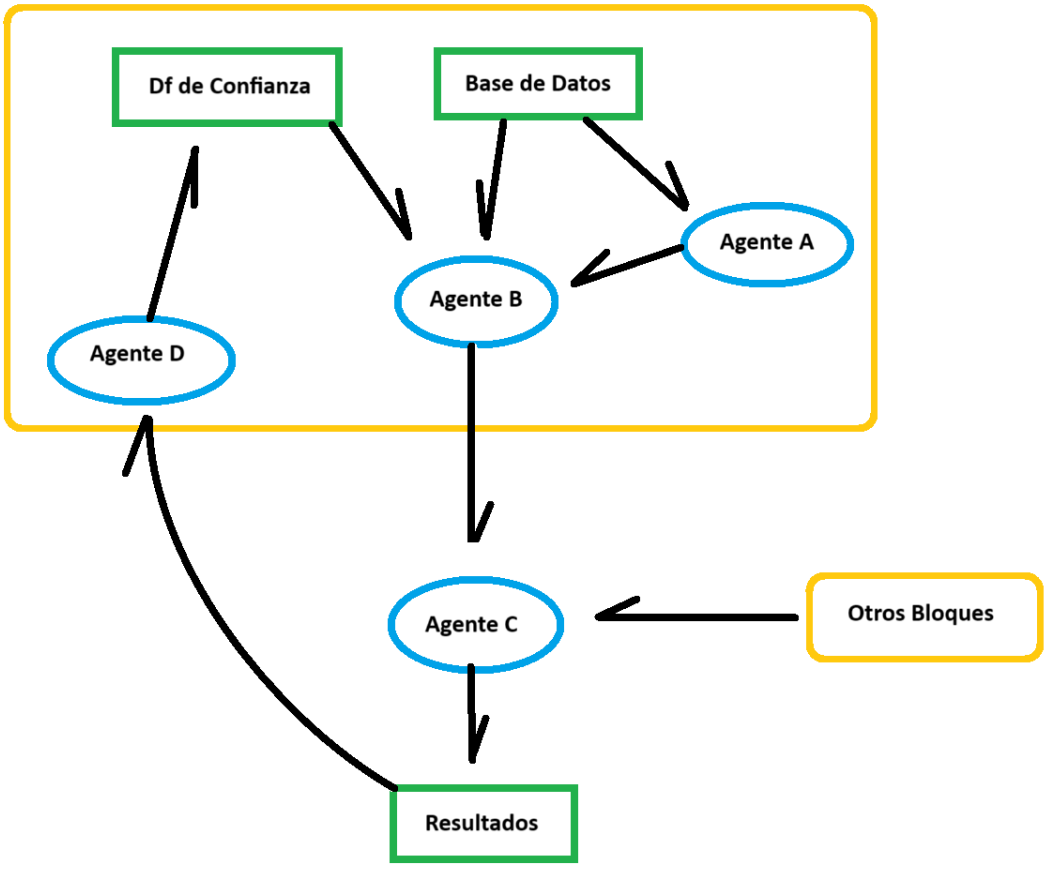
La evaluación de cada fuente, en normas generales, va a constar de tres base de datos, la principal (noticias, videos de noticieros, chats de redes sociales, etc.), una base de datos con los valores de confianza de cada ente e individuo a la hora de realizar las predicciones y una base de datos de los resultados. También dispondremos de cuatro agentes para realizar las predicciones:

Agente A: Dado unos datos de la base de datos predice en qué cantidad va a ser afectada cada tipo de cotización.

Agente B: Dado la predicción del agente A, los datos y los valores de confianza de los entes involucrados predice cuánto va a variar el valor de cada cotización individual.

Agente C: Dada la predicción de los agentes A de cada fuente disponible retorna las variaciones de cada cotización y las guarda en una base de datos de resultados.

Agente D: Dado los resultados obtenidos y la confianza de los entes involucrados actualiza estos últimos valores.



Ventajas

Este modelo posee varias ventajas:

- Escalabilidad al permitir adaptar varios tipos de fuentes al mismo proyecto.
- Permite la múltiple ejecución del código gracias a permitir complementar varios módulos.
- Al tener varios agentes permite reducir la complejidad de estos ya que parte de la información ha sido procesada previamente.
- Al tener un df de las confianzas de las fuentes permite eliminar las redundancias en futuras ejecuciones.
- Permite identificar las fuentes de información más redundantes.

Desventajas

Las desventajas que posee sería:

- Alto consumo de recursos.
- Dificultad en la adaptación de diferentes tipos de fuentes de datos.
- En el caso de usar aprendizaje automático obtenemos una dificultad en el entrenamiento.

Pese a las ventajas que brindaría en la predicción de tendencias inmediatas del mercado su alto requerimiento computacional y posibles cuellos de botella retrasarían la predicción lo suficiente para no ser rentable en este ámbito.