

Trabajos relacionados (estado del arte)

1. Introducción

El avance de las tecnologías de procesamiento del lenguaje natural (NLP) y la inteligencia artificial ha facilitado la creación de sistemas automatizados como los chatbots, capaces de interactuar con los usuarios en lenguaje natural. Estos sistemas han encontrado aplicaciones en múltiples áreas, desde atención al cliente hasta la asesoría legal. Esta tesis se enfoca en el desarrollo de un chatbot especializado en proporcionar asistencia relacionada con los derechos del consumidor.

Para lograr que un chatbot comprenda y responda de manera precisa sobre normativas legales, es necesario integrar tecnologías avanzadas. Los modelos de lenguaje a gran escala (LLMs), como GPT-3, ofrecen la capacidad de generar respuestas en lenguaje natural, mientras que técnicas como Retrieval-Augmented Generation (RAG) garantizan que las respuestas estén basadas en fuentes confiables, como leyes o documentos legales.

2. Descripción de las Tecnologías

2.1 Chatbots

Los chatbots son programas de software diseñados para interactuar con los usuarios a través del lenguaje natural. Utilizan técnicas de NLP para interpretar y generar respuestas automatizadas. Según Følstad et al. (2021), el desarrollo de chatbots debe enfocarse en mejorar la experiencia del usuario y la capacidad del sistema para adaptarse a diversos contextos (Følstad et al., 2021, pp. 603-604).

2.2 RAG

El enfoque Retrieval-Augmented Generation (RAG) combina modelos de lenguaje preentrenados con recuperación de información no paramétrica, permitiendo que los modelos generen respuestas basadas en datos externos. Esto es particularmente útil en tareas que requieren manejar grandes cantidades de datos especializados, como la interpretación de leyes (Lewis et al., 2020, pp. 1-2).

2.3 LLMs

Los Modelos de Lenguaje a Gran Escala (LLMs) son sistemas de inteligencia artificial entrenados en enormes volúmenes de datos textuales que utilizan la arquitectura Transformer. Estos modelos son capaces de realizar tareas complejas como traducción automática, generación de texto, y análisis de sentimiento. Sin embargo, no comprenden el contexto de manera similar a los humanos, sino que se basan en correlaciones estadísticas (Shanahan, 2024, pp. 68-69).

3. Evolución Histórica

3.1 Chatbots

El desarrollo de los chatbots comenzó con ELIZA en 1966, un programa que simulaba conversaciones mediante coincidencias de patrones. Følstad et al. (2021) destacan que, aunque los chatbots actuales han mejorado en la comprensión del lenguaje natural, aún enfrentan desafíos relacionados con la personalización y la fluidez en diversos contextos (Følstad et al., 2021, pp. 605-606).

3.2 RAG

Antes de RAG, enfoques como REALM y ORQA ya combinaban generación y recuperación, pero se limitaban a tareas extractivas. RAG, introducido por Lewis et al. (2020), integró modelos generativos como BART, mejorando la precisión de las respuestas en tareas de generación de texto (Lewis et al., 2020, pp. 3-4).

3.3 LLMs

El desarrollo de los LLMs fue impulsado por la arquitectura Transformer, que se popularizó con GPT-2 y GPT-3. Estos modelos han demostrado ser extremadamente efectivos para generar texto y realizar tareas complejas, aunque su capacidad de predicción de secuencias de palabras no equivale a una verdadera comprensión (Shanahan, 2024, pp. 69-70).

4. Aplicaciones Relevantes

4.1 Chatbots

Los chatbots han sido implementados en sectores como atención al cliente y educación. Un ejemplo es DoNotPay, que permite resolver disputas legales de manera automatizada, demostrando su capacidad para ofrecer asistencia en áreas como los derechos del consumidor (Adamopoulou, 2020, pp. 375-376).

4.2 RAG

El modelo RAG ha sido ampliamente utilizado en tareas como la generación de preguntas y verificación de hechos. Wang et al. (2024) destacan cómo la integración de técnicas adecuadas puede mejorar el rendimiento en tareas complejas, asegurando que el sistema maneje grandes volúmenes de información de manera eficiente (Wang et al., 2024, pp. 6-7).

4.3 LLMs

Los LLMs se utilizan en aplicaciones como generación de contenido y análisis de datos. Shanahan (2024) advierte que, aunque los LLMs pueden generar respuestas convincentes, no tienen una verdadera comprensión, lo que implica ciertos límites en su aplicabilidad para tareas que requieren conocimiento profundo (Shanahan, 2024, pp. 70-72).

5. Relevancia para el Proyecto

5.1 Chatbots

Para el chatbot de derechos del consumidor, esta tecnología permitirá que los usuarios realicen consultas legales y reciban respuestas basadas en normativas específicas. Følstad et al. (2021) subrayan la importancia de garantizar una experiencia de usuario accesible y ética (Følstad et al., 2021, pp. 610-611).

5.2 RAG

El enfoque RAG es esencial para acceder a grandes bases de datos legales y garantizar que las respuestas estén respaldadas por normativas precisas. Esto es crucial para ofrecer un servicio confiable en el dominio legal (Lewis et al., 2020, pp. 7-8).

5.3 LLMs

Los LLMs permitirán generar respuestas detalladas a preguntas legales complejas. Sin embargo, Shanahan (2024) advierte sobre la necesidad de supervisión humana para evitar que los LLMs generen respuestas erróneas (Shanahan, 2024, pp. 72-74).

Bibliografía:

Adamopoulou, E., & Moussiades, L. (2020). An Overview of Chatbot Technology. En I.

Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *Artificial Intelligence Applications and Innovations* (pp. 373-383). Springer International Publishing. https://doi.org/10.1007/978-3-030-49186-4_31

Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M.,

Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer von Wolff, R., Hobert, S., & Luger, E. (2021). Future directions for chatbot research: An interdisciplinary research agenda. *Computing*, 103(12), 2915-2942. <https://doi.org/10.1007/s00607-021-01016-7>

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M.,

Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.

https://proceedings.neurips.cc/paper_files/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024). *A Comprehensive Overview of Large Language Models* (No. arXiv:2307.06435). arXiv. <https://doi.org/10.48550/arXiv.2307.06435>

Shanahan, M. (2024). Talking about Large Language Models. *Communications of the ACM*, 67(2), 68-79. <https://doi.org/10.1145/3624724>

Wang, X., Wang, Z., Gao, X., Zhang, F., Wu, Y., Xu, Z., Shi, T., Wang, Z., Li, S., Qian, Q., Yin, R., Lv, C., Zheng, X., & Huang, X. (2024). *Searching for Best Practices in Retrieval-Augmented Generation* (No. arXiv:2407.01219). arXiv. <http://arxiv.org/abs/2407.01219>