INTELLIGENT PATENT PROCESSING: LEVERAGING RETRIEVAL-AUGMENTED GENERATION FOR ENHANCED CONSULTANT SERVICES

Femi Halgin[1]* and Ahmed Taiye Mohammedand[2]

[1]*Department of Media Technology and Computer Science, Linnaeus University, Sweden, Email: femihalgin@gmail.com*
[2]*Department of Information Science, Faculty of Art & Humanities, Linnaeus University, Sweden, Email: ahmadtaiye.mohammad@lnu.se*
(*Main presenter and corresponding author)

ABSTRACT

Introduction: In the era of digital transformation, characterized by high intelligence, high investment, and high risk, the patent process is undergoing a profound transformation (Alderucci, D., & Sicker, D ,2019). This transformation requires a multifaceted journey involving innovation, creativity, and technical expertise (Giczy, A. V., Pairolero, N. A., & Toole, A. A ,2022). Furthermore, between 2012 and 2021, Sweden emerged as the Nordic country with the highest number of recorded patent applications (Einar H. Dyvik, 2023). This surge in patent activity presents a complex landscape for consultants to navigate, requiring a deep understanding of patent law, technical expertise, and attention to detail. Innovative systems such as the Patent Office Action Response Intelligence System utilize techniques like Latent Dirichlet Allocation (LDA) and cosine similarity to analyze patent data (Chu, J. M., Lo, H. C., Hsiang, J., & Cho, C. C., 2024). However, this approach has limitations, including the need for effective text pre-processing, careful model parameter selection, and rigorous model reliability evaluation. Additionally, interpreting the resulting topics requires a deep understanding of the data and the nuances of the patent application process (Maier et al., 2021). AI has already demonstrated its potential in patent classification (Lee & Hsiang, 2020b) and claim generation (Lee & Hsiang, 2020a), showcasing its ability to automate routine tasks and reduce human work (Chikhaoui, E., & Mehar, S ,2020). However, with the advent of generative AI (GenAI), the patent process is poised to revolutionize the consultant role. As Large Language Models (LLMs) continue to evolve (Workshop et al., 2022; Touvron et al., 2023), AI-assisted agents are demonstrating remarkable abilities in understanding Natural Language Process (NLP) and manipulating well-formatted text drafted by humans (Ali, A. A. S., & Shandilya, V. K. ,2021). Despite advancements in AI-assisted patent processing, reliability concerns persist, where outputs generated may be misleading, hallucinated, or incorrect. To address this, the Retrieval-Augmented Generation (RAG) framework prioritizes transparency, explainability, and accountability to ensure trustworthy AI outputs (Li, J., Yuan, Y., & Zhang, Z, 2024). As the journey of technological advancement continues, it is essential to explore the potential of RAG in transforming the patent process.

ISBN: 978-91-8082-142-1

This study explores the transformative impact of GenAI using RAG techniques on the consultant role in patent processing, examining the benefits, challenges, and implications for consultants, clients, and the patent industry.

Challenges: Despite the automation of routine tasks in patent processing, a critical gap remains in the communication and knowledge-sharing process between examiners, inventors, and patent agents (consultation). This gap leads to inefficient communication, delays, and workload overflow, as patent agents struggle to manage multiple cases, compromising client service and case management quality. Furthermore, the lack of standardized benchmarks to evaluate Large Language Models' understanding of intellectual property-related concepts and regulations hinders the development of effective AI-assisted patent processing tools. Moreover, the unreliability of LLMs, which can generate hallucinate or incorrect outputs, raises concerns about the trustworthiness of AI-generated patent-related information, emphasizing the need for rigorous validation and ongoing refinement is needed by field and technical expert.

Research Questions: To explore the potential of GenAI using RAG in transforming the consultant role in patent processing these research questions are formulated.

RQ 1: What are the practical and research benefits and challenges of using GenAI in patent processing and consultation?

RQ 2: How to understand the knowledge and information gap of GenAI on the communication and knowledge-sharing processes between patent examiners, inventors, and patent consultants? How can RAG enhance the accuracy and effectiveness of these interactions compared with existing models?

Methods: This study will adopt a mixed methods approach to explore the integration of Generative AI (GenAI) and Retrieval-Augmented Generation (RAG) in the patent process. The methodology will unfold in three phases as depicted in the image.

Phase 1: Literature Review & Problem Identification: The study begins with an extensive analysis of the literature, focusing on acquiring knowledge related to AI's role in the patent process. This phase also includes identifying problems and gaps in current methods, followed by an examination of existing models in patent workflows.

Phase 2: Data Collection & Prototype Design: In this phase, the study will involve data collection through multiple qualitative and quantitative research methods. Case studies will examine organizations that have integrated AI in their patent processes, while surveys, questionnaires, and focus groups will capture stakeholder perceptions and group dynamics around AI tools. Interviews with industry experts will offer deeper insights into AI’s impact on the patent process, and workshops will facilitate collaborative discussions. Using the collected data, a prototype will be designed and analysed, integrating GenAI with RAG. The study will involve both thematic analysis and quantitative research methods to assess key outcomes.

Phase 3: Validation & Results: The final phase involves comparing the AI prototype to existing benchmarks to evaluate its effectiveness in enhancing patent examination and innovation. Workshops and expert interviews will validate the findings and provide feedback on the practical integration of AI tools.

Conclusion: In this proposal, the study will explore the transformative potential of GenAI with RAG in revolutionizing the role of consultants in the patent processing industry. By integrating these advanced AI technologies, the study aims to address critical challenges such as inefficiencies in communication and knowledge-sharing between patent examiners, inventors, and patent consultants. Ultimately enhancing the efficiency and accuracy of patent processing, GenAI enhanced by RAG promises to automate routine tasks, reduce human workload, and ensure reliable, transparent, and accountable AI outputs. Thereby, streamlining the patent examination process and handling complex textual data with precision. Future work will focus on developing standardized benchmarks for evaluating AI models, enhancing knowledge graphs and databases, and ensuring ethical and transparent AI practices. Additionally, pilot programs and case studies will be conducted to provide valuable insights and best practices from the real-life experiences of industry stakeholders.

Possible Deliverables: Deliverables include the development of GenAI system with RAG, a proof of concept with custom knowledge graphs and databases for patent processing.

Keywords: Patent processing, GenAI, Patent consultants, RAG

REFERENCES

Alderucci, D., & Sicker, D. (2019). Applying artificial intelligence to the patent system. Technology & Innovation, 20(4), 415-425.

Chikhaoui, E., & Mehar, S. (2020). Artificial intelligence (AI) collides with patent law. Journal of Legal, Ethical & Regulatory Issues, 23, 1.

Giczy, A. V., Pairolero, N. A., & Toole, A. A. (2022). Identifying artificial intelligence (AI) invention: A novel AI patent dataset. The Journal of Technology Transfer, 47(2), 476- 505.

BigScience Workshop, Teven Le Scao, A., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A. S., Yvon, F., et al. (2022). Bloom: A 176b-parameter open-access multilingual language model. arXiv preprint arXiv:2211.05100.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.

Lee, J.-S., & Hsiang, J. (2020a). Patent claim generation by fine-tuning openai gpt-2. World Patent Information, 62, 101983. • Lee, J.-S., & Hsiang, J. (2020b). Patent classification by fine-tuning bert language model. World Patent Information, 61, 101965.

Chu, J. M., Lo, H. C., Hsiang, J., & Cho, C. C. (2024). From PARIS to LE-PARIS: Toward patent response automation with recommender systems and collaborative large language models. arXiv preprint arXiv:2402.00421.

Manoharan, G., Razak, A., Rao, C. G., Ashtikar, S. P., & Nivedha, M. (2024). Artificial intelligence at the helm: Steering the modern business landscape toward progress. In The ethical frontier of AI and data analysis (pp. 72-99). IGI Global.

Li, J., Yuan, Y., & Zhang, Z. (2024). Enhancing LLM factual accuracy with RAG to counter hallucinations: A case study on domain-specific queries in private knowledge-bases. arXiv preprint arXiv:2403.10446.

Ali, A. A. S., & Shandilya, V. K. (2021). AI-Natural Language Processing (NLP). International Journal for Research in Applied Science and Engineering Technology, 9, 135-140.

Dyvik, E. H. (2023, December 7). Number of granted patents in Sweden 2012-2021. Nordic Statistics Database. Retrieved from https://www.statista.com/statistics/1083947/number-of-granted-patents-from-sweden/

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., ... & Adam, S. (2021). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. In Computational methods for communication science (pp. 13- 38). Routledge