Summary post

Large language models (LLMs) have the potential to improve efficiency across sectors such as legal, administrative, and creative fields. However, they also pose risks, including generating false or biased content and lacking true understanding. The complexity of these models makes errors difficult to trace and fix, emphasizing the need for responsible development through transparency, bias reduction, and ethical oversight.

Dinh Khoi Dang suggested adding real-world examples to illustrate LLM applications, such as tools like ChatLaw for legal document summarization or Eskritor for AI-generated poetry (Cui et al., 2024; AI Poem Generator: Create Beautiful Poems Effortlessly, 2025). He also recommended exploring ethical concerns in more depth by including cases of bias and harmful outputs. Martyna provided an example of Amazon's AI recruiting tool, which exhibited gender bias due to skewed training data (BBC, 2018). She proposed solutions such as penalizing biased outputs and using more diverse datasets to prevent discriminatory outcomes.

Jaafar emphasized the importance of human oversight, especially in high-risk fields like healthcare. Martyna echoed this concern, citing retinal disease diagnostic algorithms that performed worse for darker-skinned individuals (60.5%) compared to lighter-skinned individuals (73%), demonstrating the potential harm of biased training data (Burlina et al., 2020).

Dinh Khoi Dang also suggested discussing LLM-generated misinformation. A notable case involved a legal proceeding in Minnesota, where an expert used ChatGPT to compile citations, leading to non-existent sources that undermined his testimony. The presiding judge excluded the testimony, highlighting the risks of relying on LLMs in critical contexts such as law (Thomas, 2025). Similarly, Jaafar highlighted challenges in AI interpretability and compliance in regulated sectors

like law and finance, which could be addressed through Explainable AI (XAI) methods such as SHAP and LIME to improve transparency and accountability (Minh et al., 2021).

In conclusion, while LLMs present numerous benefits, responsible development and deployment are essential to mitigate their risks, ensuring ethical and equitable outcomes in critical areas such as law, healthcare, and finance.

References:

AI Poem Generator: Create Beautiful Poems Effortlessly (2025) Eskritor. Available at: https://eskritor.com/ai-poem-generator/ (Accessed: 22 January 2025).
BBC (2018) 'Amazon scrapped "sexist AI" tool', 10 October. Available at: https://www.bbc.com/news/technology-45809919 (Accessed: 22 January 2025).
Burlina, P. et al. (2020) 'Addressing Artificial Intelligence Bias in Retinal Disease Diagnostics'. arXiv. Available at: https://doi.org/10.48550/arXiv.2004.13515.
Cui, J. et al. (2024) 'Chatlaw: A Multi-Agent Collaborative Legal Assistant with Knowledge Graph Enhanced Mixture-of-Experts Large Language Model'. arXiv. Available at: https://doi.org/10.48550/arXiv.2306.16092.
Minh, D. et al. (2021) 'Explainable artificial intelligence: a comprehensive review', Artificial Intelligence Review [Preprint]. Available at: https://link.springer.com/article/10.1007/s10462-021-10088-y?utm_source=chatgpt.com (Accessed: 22 January 2025).
Thomas, D. (2025) Judge rebukes Minnesota over AI errors in 'deepfakes' lawsuit | Reuters. Available at: https://www.reuters.com/legal/government/judge-rebukes-minnesota-over-ai-errors-deepfakes-lawsuit-2025-01-13/?utm_source=chatgpt.com (Accessed: 22 January 2025).