26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Generative AI Enhancing Understanding of Vast Integrated Environmental Assessments

Ben Leighton [a], Ashlin Lee [a], Peter Baker [a]

[a] CSIRO Environment
Email: Ben.Leighton@csiro.au

Abstract: The Roper River Water Resource Assessment (RRWRA) is an integrated science-based assessment addressing the potential for water and agricultural development in the Roper River Catchment (CSIRO 2024), it consists of over 3000 pages of PDF reports. Such long and complex environmental reports can be difficult to effectively understand and use. Large Language Models (LLMs) promise to augment and improve human ability to comprehend and retrieve information and research shows LLMs are effective in comprehension, summarization and targeted question answering across inputs text (Tao 2024, Bai 2025). However, LLMs are limited in how much input text they can process and the RRWRA reports far exceed that capacity. Retrieval Augmented Generation (RAG) systems combine LLMs with semantic databases to mitigate input length limits (Lewis 2021). Prior work shows that RAG systems can perform as well as, or better than, LLMs alone, e.g. RAG systems have outperformed junior doctors on generating preoperative instructions using a collection of guideline documents (Ke 2024). We have developed the Roper River AI Assistant (RRAI), a RAG system for finding and synthesizing information across the RRWRA reports. The RRAI extends baseline RAG by adding citation fidelity: after generating a first pass answer RRAI uses an additional LLM pass to retrieve supporting quotations. Those quotations are then highlighted in a rendered display of the source PDF. Cited quotations are verified outside the LLM using exact match regular expressions to ensure their veracity. Relating answers to citations clearly visible in the PDFs provides users with a direct means of assessing answers against relevant source material. The RRAI promotes human-in-the-loop, encouraging users to verify answers and to avoid over reliance and assumptions on the accuracy of the RAG generated answer alone. In addition, RRAI uses the unstructured library (Shen 2021) to parse tables and images accurately enabling inclusion in RAG answers. It also includes prototype geocoding and mapping of relevant locations. Assessment using the Ragas framework shows average Faithfulness at 0.928, Answer Relevancy at 0.847, Context Precision at 0.884 and Answer Correctness 0.632 (Es 2025). We plan to further develop and assess the RRAI software as well as supporting social systems and collaborative practice to help address trust, accountability and responsible use (Ng and Zhang 2025, Nah 2023). Our aim is for the RRAI to be readily adapted across a wide range of environmental and water science use cases to inform decision makers and other environmental stakeholders with a richness and accuracy previously unattainable.

REFERENCES

Bai Y, Tu S, Zhang J, Peng H, Wang X, Lv X, Cao S, Xu J, Hou L, Dong Y, Tang J and Li J (2025) LongBench v2: Towards deeper understanding and reasoning on realistic long-context multitasks. arXiv preprint arXiv:2412.15204. Available at: https://arxiv.org/abs/2412.15204

CSIRO (2024) Roper River water resource assessment. CSIRO report. Available at: https://www.csiro.au/en/research/natural-environment/water/water-resource-assessment/roper-river-water-resource-assessment

Es S, James J, Espinosa-Anke L and Schockaert S (2025) Ragas: Automated evaluation of retrieval augmented generation. arXiv preprint arXiv:2309.15217. Available at: https://arxiv.org/abs/2309.15217

Ke Y, Jin L, Elangovan K, Abdullah H R, Liu N, Sia A T H, Soh C R, Tung J Y M, Ong J C L and Ting D S W (2024) Development and testing of retrieval augmented generation in large language models – a case study report. arXiv preprint arXiv:2402.01733. Available at: https://arxiv.org/abs/2402.01733

Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih W, Rocktäschel T, Riedel S and Kiela D (2021) Retrieval-augmented generation for knowledge-intensive NLP tasks. arXiv preprint arXiv:2005.11401. Available at: https://arxiv.org/abs/2005.11401

Nah F F-H, Zheng R, Cai J, Siau K and Chen L (2023) Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research 25(3), 277–304. DOI: 10.1080/15228053.2023.2233814

Ng S W T and Zhang R (2025) Trust in AI chatbots: a systematic review. Telematics and Informatics 97, 102240. DOI: 10.1016/j.tele.2025.102240

Shen Z, Zhang R, Dell M, Lee BCG, Carlson J and Li W (2021) LayoutParser: A unified toolkit for deep learning based document image analysis. Available at: https://layout-parser.github.io

Tao K, Osman Z A, Tzou P L, Rhee S Y, Ahluwalia V and Shafer R W (2024) GPT-4 performance on querying scientific publications: reproducibility, accuracy, and impact of an instruction sheet. BMC Medical Research Methodology 24, 139. DOI: 10.1186/s12874-024-02253-y

Keywords: *Generative AI, Integrated Environmental Assessments*

