

# Annotated Bibliography

Yajas Kandel

February 19, 2026

## References

- [1] Vasileios Alevizos, Nikitas Gerolimos, Eleni Aikaterini Leligkou, Giorgos Hompis, Georgios Priniotakis, and George A. Papakostas. Sustainable swarm intelligence: Assessing carbon-aware optimization in high-performance AI systems. *Technologies*, 13(10), 2025.

This 2025 peer-reviewed article from the journal *Technologies* evaluates the carbon footprint of swarm-based optimization algorithms within High-Performance Computing (HPC) environments. The authors introduce a "Emission Impact Metric" that links algorithmic behavior to real-time power consumption and grid-intensity data. By analyzing 41 different algorithms across 480 experimental runs, the study highlights a significant efficiency gap; for instance, Particle Swarm Optimization emitted only 24.9 g of  $CO_2$  per optimum compared to 61.3 g for standard Grid Search. The researchers propose a framework for auditable runtime controls that can throttle or release AI processes based on carbon objectives without sacrificing the quality of the solution. Seeing students working with Swarm Intelligence concepts for their Independent Study at Wooster got me thinking about how I could apply these concepts to something I'm passionate about. I think this is one such study.

- [2] Douglas Arner, Janos Barberis, and Ross Buckley. Fintech, regtech, and the reconceptualization of financial regulation. *Northwestern Journal of International Law and Business*, 37:373–415, 01 2017.
- [3] Marco Bronzini, Carlo Nicolini, Bruno Lepri, Andrea Passerini, and Jacopo Staiano. Glitter or gold? deriving structured insights from sustainability reports via large language models. *EPJ Data Science*, 13, 06 2024.

This 2024 article from the peer-reviewed journal *EPJ Data Science* addresses the challenge of "greenwashing" by using LLMs to convert unstructured sustainability narratives into structured, verifiable data. The researchers implement a RAG based pipeline that maps qualitative corporate statements to specific regulatory frameworks, highlighting the gap between corporate claims and actual performance data. The credibility of this source is established

through its publication in a high-impact journal specializing in data science and its use of a multi-model evaluation strategy to ensure the validity of its findings. This paper is particularly relevant to my Independent Study’s goal of ”Automating Regulatory Alignment”. It provides a technical roadmap for how to programmatically identify mandatory reporting requirements within the ESRS E1 framework. By following the authors’ approach to data structuring, I can improve the precision of my software’s ability to flag missing or non-compliant climate disclosures. Furthermore, the study’s focus on transparency aligns with my interests in Data Science and Environmental Studies, providing a clear example of how AI can be used for corporate accountability.

- [4] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms, Third Edition*. The MIT Press, 3rd edition, 2009.

This is a popular algorithms textbook which is well-cited. In particular, Part VI on graph algorithms will be of interest. Chapter 26 discusses flow networks and introduces commonly used notation. It formally describes the problem of obtaining a maximum flow and its equivalence to obtaining a minimum cut. The classical method of Ford and Fulkerson’s algorithm for finding a maximum flow is described, and it includes several examples. Additional methods for obtaining a maximum flow, including the push-relabel method, are also described. The chapter notes include additional references to specific articles which may be helpful, such as those of historical interest (the article in which an algorithm was originally proposed) as well as state-of-the-art improvements (more recent articles to improve the approach).

- [5] Kerstin Forster, Victor Wagner, Lucas Elias Keil, Maximilian A. Müller, Thorsten Sellhorn, and Stefan Feuerriegel. Tracking esg disclosures of european companies with retrieval-augmented generation. In *ICLR 2025 Workshop on Tackling Climate Change with Machine Learning*, 2025.

This 2025 peer-reviewed paper, presented at the International Conference on Learning Representations (ICLR), explores the efficacy of RAG in automating the extraction of Environmental, Social, and Governance (ESG) indicators from large-scale corporate reports. The authors develop a specialized RAG pipeline designed to handle the high linguistic complexity and lack of standardization often found in European sustainability disclosures. By utilizing a vector database to retrieve specific context before generation, the researchers demonstrate a significant reduction in model hallucinations compared to standard prompting of LLMs. This source is highly credible as it originates from a premier AI conference and utilizes a rigorous benchmarking dataset of over 600 European corporations. For my Independent Study, this paper provides a direct technical precedent for navigating ESRS E1 climate disclosures. It offers specific architectural insights into how to structure retrieval queries to capture the granular regulatory requirements mandated by European standards, which will be central to my software implementation.

- [6] Sarah Hug and Mark McKay. Problematizing AI literacy access - understanding student AI literacy from student voices. In *Proceedings of the 2025 Conference on Research on Equitable and Sustained Participation in Engineering, Computing, and Technology*, RESPECT 2025, page 339–342, New York, NY, USA, 2025. Association for Computing Machinery.
- [7] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks, 2021.
- [8] Tri-Hai Nguyen and Jason J. Jung. Swarm intelligence-based green optimization framework for sustainable transportation. *Sustainable Cities and Society*, 71:102947, 2021.

This 2021 peer-reviewed article from the journal *Sustainable Cities and Society* presents a decentralized green traffic optimization framework aimed at creating sustainable transportation systems within smart cities. The core innovation of the paper is the concept of a "repelling pheromone," which, unlike traditional pheromones that attract agents to a single path, generates a repulsive force to distribute vehicle travel paths across a road network. Validated through the Simulation of Urban Mobility (SUMO) platform, the study demonstrates significant environmental benefits, including a 13–19% reduction in fuel consumption and emissions. This study provides a technical roadmap for how swarm-based IoT systems can directly mitigate the "Tragedy of the Commons" in urban infrastructure, which is a highly relevant issue to address.

- [9] Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. Retrieval augmentation reduces hallucination in conversation, 2021.
- [10] Aparna Vinayan Kozhipuram, Samar Shailendra, and Rajan Kadel. Retrieval-augmented generation vs. baseline llms: A multi-metric evaluation for knowledge-intensive content. *Information*, 16(9), 2025.

This paper compared four LLMs (from 1.1 Billion parameters to 13 Billion parameters), to see if Retrieval Augmented Generation (RAG) helps smaller models catch up to big ones. The models used were TinyLlama 1.1B, Mistral 7B, Llama 3.1 8B and Llama 1 13 B. RAG is an AI architecture that optimizes the performance of models by allowing them to access external knowledge bases, enabling them to generate responses based on real, up-to-date data rather than solely relying on their training data. Unlike LLMs that rely on static training data, RAG pulls relevant text from databases, uploaded documents, or web sources. The researchers used a "multi -metric" approach using seven lexical and semantic scores to evaluate the performance gains of RAG LLM's over baseline LLMs. This is a 2025 peer reviewed article from the *Information Journal*, published by MDPI. Since I am interested in Environmental Studies and Data Science, one area of research for me could be

looking into how RAG models might be more sustainable or cheaper to run compared to massive models.