

Annotated Bibliography

Energy Efficient AI & Ethical Studies

Cadler D'ahiti

University of Wisconsin–Milwaukee, McNair Scholars Program

Source 1

The Art of Designing Socially Intelligent Agents: Science, Fiction, and the Human in the Loop

Key Questions

1. What is the overarching theme of the article? The article develops a theoretical framework for designing “socially intelligent agents” (SIAs) by grounding agent design in human social psychology, including storytelling, empathy, embodiment, and historical/ecological grounding, rather than treating sociality as a surface-level, imitative behavior.

2. What is the behavior under investigation? There is no directly observed behavior; the article instead examines the conceptual relationship between human social cognition and the design choices that make an artificial agent appear (or actually be) socially intelligent, using illustrative examples such as cyberpet technology.

3. What is the data source, and what are the findings derived from using these data? There is no empirical dataset. The author draws on prior theoretical literature (e.g., Turing, Franklin & Graesser, Luck & d'Inverno) and illustrative examples to construct a terminology (embodied, autonomous, believable, and autobiographic agents) and a list of design guidelines for SIA technology.

4. What is missing from the article? The article does not test its guidelines against real users or systems, and, writing in 1998, it does not anticipate the computational or energy costs of building such agents at scale, a gap that later, data-driven sources in this bibliography address directly.

5. What are its strengths and weaknesses? Strengths: foundational and interdisciplinary, still widely cited (161+ citations) in later robotics and agent-design literature. Weaknesses: purely conceptual with no empirical validation, and dated relative to generative-AI-era agents.

Complete Citation

Dautenhahn, K. (1998). The art of designing socially intelligent agents: Science, fiction, and the human in the loop. *Applied Artificial Intelligence*, 12(7–8), 573–617.

<https://doi.org/10.1080/088395198117550>

Annotation

The purpose of this article is to propose a cognitive-technology approach to designing socially intelligent agents that is grounded in human social psychology rather than surface-level mimicry. Its guiding hypothesis is that grounding SIA design in concepts such as embodiment, storytelling, and empathy will produce more believable and cognitively well-balanced agents than approaches that merely exploit humans' tendency to anthropomorphize. The study draws on none empirical; a synthesis of prior theoretical literature and illustrative examples (e.g., cyberpet technology). Key findings show that the article proposes a terminology for SIA research and a list of general design guidelines (e.g., humans as embodied, active, and individual agents) intended to guide future SIA development. Regarding whether these findings supported the author's hypothesis, because no empirical study was conducted, the article's claims are argued and illustrated rather than tested; support for the thesis is conceptual, not data-driven.

Key Terms: socially intelligent agents (SIA), embodied artificial life (EAL), cognitive technology, believability, autonomous agents, anthropomorphism, autobiographic agents

Source 2

Sparking Curiosity with Applied Intelligence

Key Questions

1. What is the overarching theme of the article? The article argues that K–12 education should shift students from passively “using” generative AI for quick answers toward “applying” AI as a collaborative partner to solve messy, real-world problems, a shift the authors call applied intelligence.

2. What is the behavior under investigation? The behavior of interest is how students and teachers engage with AI tools in the classroom: whether that engagement is passive (offloading cognitive work to AI) or active (designing, constraining, and collaborating with AI systems).

3. What is the data source, and what are the findings derived from using these data? The article is not a controlled study; it draws on practitioner case examples (a student-built oral-cancer screening app, an ice-thickness safety device, a student-designed AI writing tutor, a high schooler's AI-assisted astronomy discovery) and cites one external labor-market study (Brynjolfsson et al., 2025) showing employment declines concentrated in AI-exposed entry-level occupations.

4. What is missing from the article? The article offers no systematic data on learning outcomes comparing applied versus passive AI use, no data on the computational or energy costs of the student AI projects it describes, and relies on anecdote rather than a defined study sample.

5. What are its strengths and weaknesses? Strengths: timely, practice-oriented, and grounded in concrete classroom examples plus one credible labor-market citation. Weaknesses: anecdotal and non-empirical; it does not quantify the learning or resource-efficiency claims it makes.

Complete Citation

Radday, E. A., & Mervis, M. (2026, April). Sparking curiosity with applied intelligence. *Educational Leadership*, 23–27.

Annotation

The purpose of this article is to argue that K–12 education should teach applied intelligence (using AI as a collaborative partner on real problems) rather than teaching AI merely as an answer-generating tool. Its guiding hypothesis is that shifting students from passive AI use to active, applied AI use will foster critical thinking, ethics, and system-design skills that the labor market increasingly rewards. The study draws on illustrative classroom case studies and one cited external working paper (Brynjolfsson, Chandar, & Chen, 2025, Stanford Digital Economy Lab). Key findings show that case examples show that students who design and constrain AI systems (rather than simply query them) develop deeper engagement and skills; the cited labor data show early-career employment declines concentrated in AI-automatable occupations. Regarding whether these findings supported the author's hypothesis, the anecdotal cases and the cited labor statistic are consistent with the article's argument, but because no controlled comparison was run, the support is illustrative rather than experimentally confirmed.

Key Terms: applied intelligence, generative AI in education, critical thinking, vibe coding, AI literacy, human-AI collaboration

Source 3

Conversational Agent with Integrated Tracking Based on Generative AI Considering the Effect of Document Order and Structure

Key Questions

1. What is the overarching theme of the article? The article examines whether the order and structure of the documents used to train a generative-AI conversational agent affects the quality of the agent's responses, in the context of an academic-procedure tracking assistant.

2. What is the behavior under investigation? The behavior under investigation is the response-generation quality (accuracy, coherence, and semantic relevance) of a GAI-based conversational agent when trained on three structurally different versions of the same source document.

3. What is the data source, and what are the findings derived from using these data? The data source was three versions (Ver1, Ver2, Ver3) of one Spanish-language university document on social-service procedures, progressively cleaned and restructured (irrelevant content removed, tables clarified, images converted to text, headings clarified). Each version was tested with 10 standardized questions and scored on six NLP metrics (ROUGE-1, ROUGE-2, ROUGE-L, METEOR, BERTScore, Sem-NCG) plus expert qualitative review. Ver3 produced the strongest scores on most advanced (semantic) metrics, confirming that better document structure improves response quality.

4. What is missing from the article? The study is limited to a single base document and domain (university social-service procedures) at one institution, which restricts generalizability, and it does not measure whether better-structured documents also reduce the agent's computational or energy cost, a gap this study addresses directly.

5. What are its strengths and weaknesses? Strengths: a controlled, comparative design combining six quantitative metrics with expert qualitative review. Weaknesses: narrow domain and single-document scope, and mixed results on some metrics (Ver2 slightly outperformed Ver3 on exact-keyword-overlap metrics).

Complete Citation

Gutiérrez Tafoya, R. R., Quintana López, M., López Chau, A., & Morales Escobar, S. J. (2026). Conversational agent with integrated tracking based on generative AI considering the effect of document order and structure. *International Journal of Combinatorial Optimization Problems and Informatics*, 17(2), 190–205.
<https://doi.org/10.61467/2007.1558.2026.v17i2.1267>

Annotation

The purpose of this article is to determine whether the order and structure of training data affect the quality of responses generated by a GAI conversational agent. Its guiding hypothesis is that a more optimized document structure (Ver3) would produce measurably better agent responses than less-structured versions (Ver1, Ver2). The study draws on three restructured versions of one Spanish-language administrative document, tested with 10 standardized prompts and evaluated using ROUGE-1/2/L, METEOR, BERTScore, and Sem-NCG, plus expert qualitative analysis. Key findings show that Ver3 scored highest on most advanced semantic/structural metrics (ROUGE-2, ROUGE-L, BERTScore, Sem-NCG) and Ver1 consistently scored lowest, while Ver2 slightly outperformed on exact-keyword metrics (ROUGE-1, METEOR). Regarding whether these findings supported the author's hypothesis, largely yes: the overall pattern of

results supports the hypothesis that better-structured input data improves response quality, with a minor exception on keyword-overlap metrics.

Key Terms: conversational agent, generative artificial intelligence (GAI), document structure, ROUGE, METEOR, BERTScore, Sem-NCG, Dialogflow CX

Source 4

Agentic and Generative AI for Autonomous Energy Systems: Reference Architecture, Open Challenges, and Research Agenda

Key Questions

1. What is the overarching theme of the article? The article proposes a reference architecture that integrates generative AI, agentic AI, and digital twins to move power grids toward safe, bounded autonomy, and it maps the open research challenges that stand in the way.

2. What is the behavior under investigation? No behavior is directly observed; this is a conceptual/architectural review examining how AI agents could perceive, reason, plan, and act within power-grid operations (e.g., self-healing grids, decentralized energy markets, digital-twin training).

3. What is the data source, and what are the findings derived from using these data? There is no primary dataset; the paper synthesizes roughly 47 existing studies into a unified architecture. It finds that generative AI (scenario synthesis, forecasting) and agentic AI (bounded decision-making) play complementary roles, and that credible autonomy in energy systems must remain supervised and bounded due to safety-critical constraints.

4. What is missing from the article? The paper does not empirically test or deploy the architecture it proposes, and, despite being about energy systems, it does not quantify the computational or energy footprint of running the AI agents it describes.

5. What are its strengths and weaknesses? Strengths: a comprehensive, well-organized synthesis connecting agentic-AI theory to a high-stakes application domain, with explicit attention to safety and human oversight. Weaknesses: purely conceptual with no original empirical data, and too recent to have accumulated independent validation.

Complete Citation

Hinov, N. (2026). Agentic and generative AI for autonomous energy systems: Reference architecture, open challenges, and research agenda. *AI*, 7(5), 176.
<https://doi.org/10.3390/ai7050176>

Annotation

The purpose of this article is to propose an integrated reference architecture for “Autonomous Energy Systems” combining generative AI, agentic AI, and digital twins, and to outline the open research challenges that remain. Its guiding hypothesis is that treating generative AI, agentic AI, and digital twins as complementary layers of one architecture, rather than isolated research tracks, produces a more coherent and deployable path toward safe autonomous energy management. The study draws on a literature synthesis (no primary empirical dataset), organized around representative workflows such as self-healing grid restoration and digital-twin-based agent training. Key findings show that generative AI and agentic AI were shown to play complementary, not competing, roles, and full autonomy was found not to be currently feasible: only bounded, human-supervised autonomy is realistic given unresolved safety, verification, and

governance gaps. Regarding whether these findings supported the author's hypothesis, yes, in the sense that the architectural synthesis is internally consistent and supports the paper's conclusion that bounded integration is more viable than isolated approaches, though this is an argued conclusion rather than an empirically tested one.

Key Terms: agentic AI, generative AI, autonomous energy systems, digital twins, self-healing grids, bounded autonomy, multi-agent reinforcement learning, smart grids

Source 5

Accelerating AI Agent Application

Key Questions

1. What is the overarching theme of the article? The article describes China's national policy push and industrial strategy to accelerate the development and adoption of AI agents, contrasting it with U.S. and European approaches.

2. What is the behavior under investigation? Rather than individual behavior, the article examines national and industrial-level adoption patterns: how Chinese government policy, technology firms, and industries are integrating AI agents into finance, power-grid dispatch, and industrial Internet platforms.

3. What is the data source, and what are the findings derived from using these data? This is a journalistic/opinion piece with no original dataset; it draws on a Chinese government action plan (Ministry of Industry and Information Technology, issued January 6, 2026) and general industry examples. It reports that China is moving from proof-of-concept to practical application

of AI agents, aided by data resources, application scenarios, and state support, but constrained by dependence on foreign AI chips.

4. What is missing from the article? The article provides no quantitative or comparative data to support its claims (e.g., that AI agents make decisions “far exceeding human efficiency”), and it does not address the environmental or energy costs of large-scale agent deployment.

5. What are its strengths and weaknesses? Strengths: useful, current geopolitical and policy context, and it supplies the “passive response” framing this study cites in its own introduction. Weaknesses: an opinion/news piece rather than peer-reviewed research, with unsupported claims and a single-country perspective.

Complete Citation

Lan, X. (2026, February 19). Accelerating AI agent application. *Beijing Review*, 34.

Annotation

The purpose of this article is to describe and explain China's strategic push to accelerate AI agent development and application across industry. Its guiding hypothesis is that China's combination of abundant data resources, wide application scenarios, and strong state strategic support gives it a distinctive advantage in AI agent development compared to the U.S. or European models. The study draws on government policy documents (the MIIT action plan) and general industry examples; no quantitative dataset. Key findings show that AI agents are described as evolving from passive tools into autonomous “partners”; China's agent development is advancing in financial and energy-grid applications but remains constrained by chip and core-algorithm dependency on foreign technology. Regarding whether these findings supported the author's

hypothesis, the article's argument is self-supporting and asserted rather than tested against independent data, so support for its claims is rhetorical rather than empirical.

Key Terms: AI agents, agentic AI, industrial Internet, passive response tools, national AI strategy, power grid dispatch

Source 6

Why AI Uses So Much Energy, and What We Can Do About It

Key Questions

1. What is the overarching theme of the article? This Q&A explains why AI, especially large language model training, consumes so much energy and outlines strategies the tech industry and research institutions can pursue to make AI more sustainable.

2. What is the behavior under investigation? The behavior examined is institutional and industry-level: the resource-consumption patterns behind GPU/TPU-based model training, cooling, and data-center operation.

3. What is the data source, and what are the findings derived from using these data? The piece is an expert interview citing the 2024 United States Data Center Energy Usage Report (Lawrence Berkeley National Laboratory). It reports that data centers used 4.4% of U.S. electricity in 2023 (a figure projected to triple by 2028), that GPU-accelerated AI server energy use grew from under 2 TWh (2017) to more than 40 TWh (2023), and that data centers could reach 20% of global electricity use by 2030–2035.

4. What is missing from the article? The article does not discuss how input or prompt quality and linguistic structure affect energy use, the specific gap this study investigates, and it does not quantify energy cost at the level of an individual query or interaction.

5. What are its strengths and weaknesses? Strengths: expert-sourced, cites a credible primary energy-usage report, and is solutions-oriented. Weaknesses: a general-audience Q&A rather than original peer-reviewed research, and broad/qualitative rather than quantitatively precise at the individual-query level.

Complete Citation

Kandemir, M. (2025, April 8). Why AI uses so much energy, and what we can do about it. *Penn State Institute of Energy and the Environment*. <https://iee.psu.edu/news/blog/why-ai-uses-so-much-energy-and-what-we-can-do-about-it>

Annotation

The purpose of this article is to explain the drivers of AI's growing energy and environmental footprint and to outline strategies for making AI development more sustainable. Its guiding hypothesis is that proactive design and infrastructure choices, including smaller domain-specific models, more efficient hardware, renewable-powered data centers, and interdisciplinary collaboration, can meaningfully reduce AI's environmental footprint without halting AI progress. The study draws on an expert interview citing the 2024 United States Data Center Energy Usage Report (Lawrence Berkeley National Laboratory) and related industry figures. Key findings show that AI training and inference are described as among the most resource-intensive computing tasks in the modern era, with data-center electricity consumption rising sharply and projected to continue growing steeply through 2030–2035. Regarding whether these findings

supported the author's hypothesis, the recommended mitigations are presented as plausible expert opinion rather than validated against outcome data; this is a solutions-oriented interview, not a controlled study.

Key Terms: data center energy usage, LLM training, GPUs/TPUs, sustainable AI, domain-specific models, renewable energy, carbon footprint

Source 7

What Is an AI Data Center?

Key Questions

1. What is the overarching theme of the article? This reference article explains the technical infrastructure, including compute, storage, networking, and power/cooling, that distinguishes AI-ready data centers from traditional data centers, and describes growth trends in AI infrastructure demand.

2. What is the behavior under investigation? Not behavioral; it is an explainer of infrastructure design trends among cloud providers and technology companies building AI data centers.

3. What is the data source, and what are the findings derived from using these data? The article aggregates third-party industry reports (Goldman Sachs, McKinsey, IBM Institute for Business Value). It reports that AI is projected to drive a 165% increase in data-center electricity demand by 2030, that global data-center capacity demand may reach 171–219 GW versus a current ~60 GW, and that AI-ready centers require GPUs/TPUs/NPUs, high-bandwidth memory, and liquid cooling.

4. What is missing from the article? The article contains no primary research or original methodology; it aggregates others' statistics, and it addresses only hardware/infrastructure factors in energy use, not software- or prompt-level efficiency.

5. What are its strengths and weaknesses? Strengths: a clear technical explainer well-sourced to credible market-research firms, useful for grounding data-center vocabulary. Weaknesses: vendor-published (IBM) content, not peer-reviewed, and occasionally promotional of IBM's own technology.

Complete Citation

Jonker, A., & Gomstyn, A. (n.d.). What is an AI data center? *IBM Think*. Retrieved July 7, 2026, from <https://www.ibm.com/think/topics/ai-data-center>

Annotation

The purpose of this article is to define what distinguishes AI-ready data centers from traditional data centers and to describe growth trends in AI infrastructure demand. Its guiding hypothesis is not applicable; this is a descriptive/reference article rather than a hypothesis-driven study. The study draws on secondary industry reports (Goldman Sachs, McKinsey, and an IBM Institute for Business Value survey). Key findings show that AI-driven electricity demand for data centers is projected to grow steeply (up to a 165% increase by 2030), and AI-ready infrastructure differs substantially from traditional data centers in compute, storage, networking, and cooling design. Regarding whether these findings supported the author's hypothesis, not applicable; descriptive/reference content rather than a tested hypothesis.

Key Terms: AI data center, hyperscale, colocation, GPU/TPU/NPU, high-bandwidth memory, liquid cooling, power usage effectiveness (PUE)

Source 8

Maximizing Renewable Energy Use in Datacenters through Watt-bit Collaboration

Key Questions

1. What is the overarching theme of the article? The article presents NTT's “watt-bit collaboration” framework for aligning datacenter power demand with renewable energy (RE) supply in Japan by integrating energy (watts) and information/communication (bits) infrastructure.

2. What is the behavior under investigation? The behavior under investigation is datacenter workload-shifting and siting behavior: how DC operations can be dynamically adjusted, through distributed workload control and DC placement, to better match fluctuating renewable-energy availability.

3. What is the data source, and what are the findings derived from using these data? The data source is Japanese government energy-mix statistics (METI) and NTT's internal engineering framework. Renewable-energy output curtailment in Japan exceeded 2 billion kWh in fiscal year 2024, concentrated in regions (Kyushu, Tohoku) with high RE potential but low datacenter demand, illustrating the geographic and temporal mismatch the paper addresses.

4. What is missing from the article? Both proposed approaches (dynamic workload control and DC placement optimization) are still conceptual or in development; the article reports no field-

deployment results or quantified energy savings yet, and it does not address AI model- or software-level efficiency, only macro-level datacenter siting and power routing.

5. What are its strengths and weaknesses? Strengths: grounded in real national energy statistics and a concrete engineering framework backed by government/industry policy.

Weaknesses: proposal-stage rather than validated results, and the Japan-specific context may limit generalizability to other electric grids.

Complete Citation

Mahmood, F., Suwabe, M., Nakanishi, K., Morisawa, Y., & Hayashi, T. (2025). Maximizing renewable energy use in datacenters through watt-bit collaboration. *NTT Technical Review*, 23(12), 26–32.

Annotation

The purpose of this article is to explain NTT's watt-bit collaboration framework for maximizing renewable energy utilization in datacenters amid AI-driven demand growth. Its guiding hypothesis is that dynamically shifting datacenter workloads and optimizing datacenter placement in response to renewable-energy availability can reduce the geographic and temporal mismatch currently causing renewable-energy curtailment. The study draws on Japanese government energy statistics (METI) on renewable-energy curtailment trends (2018–2025) and NTT's internal watt-bit collaboration engineering and policy framework. Key findings show that renewable-energy curtailment in Japan exceeded 2 billion kWh in fiscal year 2024 due to geographic and temporal mismatch between supply and datacenter demand; two proposed approaches are outlined as solutions, though both remain in development. Regarding whether these findings supported the author's hypothesis, partially: the underlying mismatch/curtailment

problem is data-supported, but the proposed solutions have not yet been empirically validated with deployment results.

Key Terms: watt-bit collaboration, renewable energy curtailment, datacenter workload distribution, DC placement optimization, GX2040 Vision, grid mismatch

Source 9

Reducing Datacenter Compute Carbon Footprint by Harnessing the Power of Specialization: Principles, Metrics, Challenges and Opportunities

Key Questions

1. What is the overarching theme of the article? The article analyzes how hardware specialization can reduce datacenter compute-related carbon emissions, and the new challenges specialization introduces across the software, system-design, and operations life cycle.

2. What is the behavior under investigation? Not a behavioral study; it is a position/analysis paper examining engineering and design trade-offs in datacenter hardware specialization (chiplets, accelerators) as they relate to embodied and operational emissions.

3. What is the data source, and what are the findings derived from using these data? The data source is aggregated global datacenter electricity statistics (460 TWh in 2022, up from about 220 TWh in 2018, projected to exceed 620 TWh by 2026) and prior emissions-quantification studies, including Strubell et al.'s estimate that training a single large language model can equal the lifetime carbon emissions of five cars. The paper finds that specialization reduces operational energy but raises embodied-emissions and system-complexity challenges.

4. What is missing from the article? The article provides no new empirical measurements of its own (it is an analytical framework paper) and does not address how software- or prompt-level inefficiency (the focus of this study) interacts with hardware specialization.

5. What are its strengths and weaknesses? Strengths: authored by a strong multi-institutional team (IBM Research and Columbia University) and published in a peer-reviewed IEEE journal; it addresses the full life cycle rather than only operational energy. Weaknesses: conceptual/framework-level rather than empirical, with no original quantitative testing.

Complete Citation

Eilam, T., Bose, P., Carloni, L. P., Cidon, A., Franke, H., Kim, M. A., Lee, E. K., Naghshineh, M., Parida, P., Stein, C. S., & Tantawi, A. N. (2024). Reducing datacenter compute carbon footprint by harnessing the power of specialization: Principles, metrics, challenges and opportunities. *IEEE Transactions on Semiconductor Manufacturing*, 37(4), 481–488.
<https://doi.org/10.1109/TSM.2024.3434331>

Annotation

The purpose of this article is to analyze the trade-offs and challenges that hardware specialization introduces for reducing datacenter carbon footprint, and to propose a unified framework for evaluating those trade-offs across a compute system's life cycle. Its guiding hypothesis is that specialization is necessary but insufficient on its own for sustainable computing, and that its downstream challenges (embodied emissions, operational complexity) must be addressed across the full life cycle to realize its full carbon-reduction potential. The study draws on aggregated global datacenter energy statistics and prior model-training emissions studies (e.g., Strubell et al.); no new primary experimental dataset. Key findings show that global datacenter electricity

use has more than doubled since 2018 and continues rising steeply; specialization reduces operational energy but raises embodied-emissions and system-complexity challenges across software, system design, and operations. Regarding whether these findings supported the author's hypothesis, yes, in the sense that the analysis consistently shows specialization's benefits are real but incomplete without addressing life-cycle trade-offs, though this is an argued/analytical conclusion rather than an experimentally tested one.

Key Terms: hardware specialization, embodied emissions, operational emissions, carbon footprint, life-cycle efficiency, chiplet design, sustainable computing

Source 10

Distribution Slack Allocation Algorithm for Energy-Aware Task Scheduling in Cloud

Datacenters

Five Key Questions

1. What is the overarching theme of the article?

The article investigates methods for reducing energy consumption in cloud data centers through an energy-aware task scheduling algorithm. The authors aim to improve the efficiency of cloud computing by minimizing energy usage while maintaining acceptable processing speed (makespan) and throughput.

2. What is the behavior under investigation?

The study examines how computational tasks are scheduled across processors within cloud data centers. Specifically, it investigates whether distributing task slack time and applying Dynamic

Voltage and Frequency Scaling (DVFS), clustering, and task duplication can reduce processor energy consumption while maintaining system performance.

3. What is the data source and what are the findings derived from using these data?

Data Source:

The researchers used simulated cloud computing environments built with the CloudSim toolkit. Their experiments consisted of randomly generated Directed Acyclic Graphs (DAGs), five simulated data centers, 200 virtual machines, and multiple processor configurations.

Findings:

- Reduced energy consumption by up to 58.3%.
- Achieved 61.2% energy savings in larger test cases.
- Improved makespan and throughput while satisfying energy constraints better than competing algorithms.

4. What is missing from the article?

The article focuses primarily on technical optimization. It does not discuss ethical implications of AI, environmental impacts beyond electricity savings, human behavior, or educational effects.

5. What are its strengths and weaknesses?

Strengths:

Detailed mathematical models, extensive simulation experiments, comparisons with existing scheduling algorithms, and they demonstrated improvements in energy efficiency

Weaknesses:

- Based on simulations rather than real-world data centers
- Highly technical and difficult for non-specialists
- Limited discussion of broader societal implications

Complete APA Citation

Berenjian, G., Motameni, H., Golsorkhtabaramiri, M., & Ebrahimnejad, A. (2021). Distribution slack allocation algorithm for energy aware task scheduling in cloud datacenters. *Journal of Intelligent & Fuzzy Systems*, 41(1), 251–272. <https://doi.org/10.3233/JIFS-201696>

Annotated Bibliography

The purpose of this study was to develop an energy-aware scheduling algorithm that reduces energy consumption in cloud data centers without significantly decreasing system performance. The authors hypothesized that combining task duplication, clustering, and Dynamic Voltage and Frequency Scaling (DVFS) with distributed slack allocation would outperform previous scheduling algorithms. Using CloudSim simulations and randomly generated DAGs, the researchers found that their EATSDCDD algorithm achieved up to 58.3% energy savings and 61.2% savings in larger task sets while maintaining throughput and reducing makespan. These findings supported the authors' hypothesis and demonstrate that improved scheduling techniques can significantly reduce data center energy consumption.

Key Terms: Cloud Computing, Cloud Data Centers, Energy Consumption, Energy Efficiency, Dynamic Voltage and Frequency Scaling (DVFS), Task Scheduling, Directed Acyclic Graph (DAG), Makespan, Throughput, Slack Time Distribution, Green Computing, Processor Optimization, Virtual Machines, CloudSim Simulation

Source 11

Telio: A Research Agent for Energy-Efficient AI and Ethical Studies

Five Key Questions

1. What is the overarching theme of the source?

The source introduces ‘**Telio**’, an AI research instrument developed to investigate the relationship between prompt quality, large language model (LLM) behavior, token usage, and estimated energy consumption. Telio is designed to support research on ethical AI use by providing measurable indicators of prompt quality, computational cost, and estimated environmental impact.

2. What is the behavior under investigation?

Telio investigates how variations in the linguistic quality of user prompts influence AI responses, token usage, and estimated energy consumption. The system evaluates prompts based on readability, redundancy, and ambiguity before submitting them to a large language model, allowing researchers to examine whether poorly structured prompts require greater computational resources.

3. What is the data source and what are the findings derived from using these data?

Data Source

The data source consists of user-generated prompts processed through the Telio application. For each prompt, the system records:

- Linguistic quality score

- Token usage
- Estimated energy consumption
- Correlation statistics between prompt quality and computational cost

Findings

Pending; results will be added after data collection.

4. What is missing from the source?

The source serves as documentation for the research instrument rather than a completed research article. As a result, it does not yet include:

- Final experimental results
- Statistical analyses
- Hypothesis testing outcomes
- Discussion of observed trends
- Conclusions based on collected data

These elements will be produced after data collection is completed.

5. What are its strengths and weaknesses?

Strengths

- Developed specifically for this research project.
- Produces transparent and reproducible measurements.
- Measures multiple dimensions of prompt quality.
- Integrates computational cost with estimated energy consumption.
- Can be replicated by future researchers using the same methodology.

Weaknesses

- Energy estimates rely on published conversion factors rather than direct hardware measurements.
- Token usage serves as a proxy for computational effort rather than a direct measure.
- Prompt quality is operationalized using heuristic scoring methods rather than universally accepted linguistic metrics.
- The instrument depends on external large language model APIs.

Complete APA Citation

D'ahiti, C. (2026). Telio: A research instrument for energy-efficient AI and ethical studies (Version 1.0) [Computer software].

Annotated Bibliography

‘Telio’ was developed to investigate the relationship between prompt quality, AI behavior, token usage, and estimated energy consumption during interactions with large language models. The purpose of the software is to provide a reproducible research instrument capable of evaluating linguistic characteristics of prompts while estimating the computational resources required to generate AI responses. The underlying hypothesis is that prompts with lower linguistic quality may require greater computational effort, resulting in higher token usage and increased estimated energy consumption. The instrument collects data by scoring prompt readability, lexical redundancy, and ambiguity before recording token usage and estimating energy costs using published conversion factors. The agent’s design supports the research objective of examining ethical AI use while exploring opportunities to improve energy efficiency and encourage more effective human–AI interaction.

Key Terms

- ‘Telio’
- Artificial Intelligence
- Large Language Models (LLMs)
- Prompt Engineering
- Prompt Quality
- Token Usage
- Energy Consumption
- Computational Cost
- Linguistic Quality
- Readability
- Lexical Redundancy
- Ambiguity
- Ethical AI
- Human–AI Interaction
- AI Agents
- Sustainable Computing
- Research Instrument
- Energy Efficiency