Proposed Solution Report

Design and Creative Technologies

Torrens University, Australia

Student: Luis Guilherme de Barros Andrade Faria - A00187785

Subject Code: HCD 402

Subject Name: Research Methodologies

Assessment No.: 2

Title of Assessment: Proposed Solution Report

Lecturer: Dr. Omid Haas

Date: Nov 2025

Table of Contents

1. Introduction

It is impossible not to hear about **AI agents** these days. They dominate headlines, warning that Artificial Intelligence is replacing jobs, flood LinkedIn posts offering "step-by-step" tutorials ("Comment '*AGENT*' to get the workflow"), and power popular automation tools like **n8n**, **Claude**, **Temporal**, **Motion**, **Boomi**, **Copilot**, etc. The hype is real – and anyone who works with technology is already part of it.

As illustrated in *Figure 1*, global search interest for the term "**AI agent**" has skyrocketed between January 2023 and October 2025, confirming the acceleration of this new technological wave. *Figure 2* further highlights related topics and queries, reflecting how quickly this concept has entered mainstream technical vocabulary.

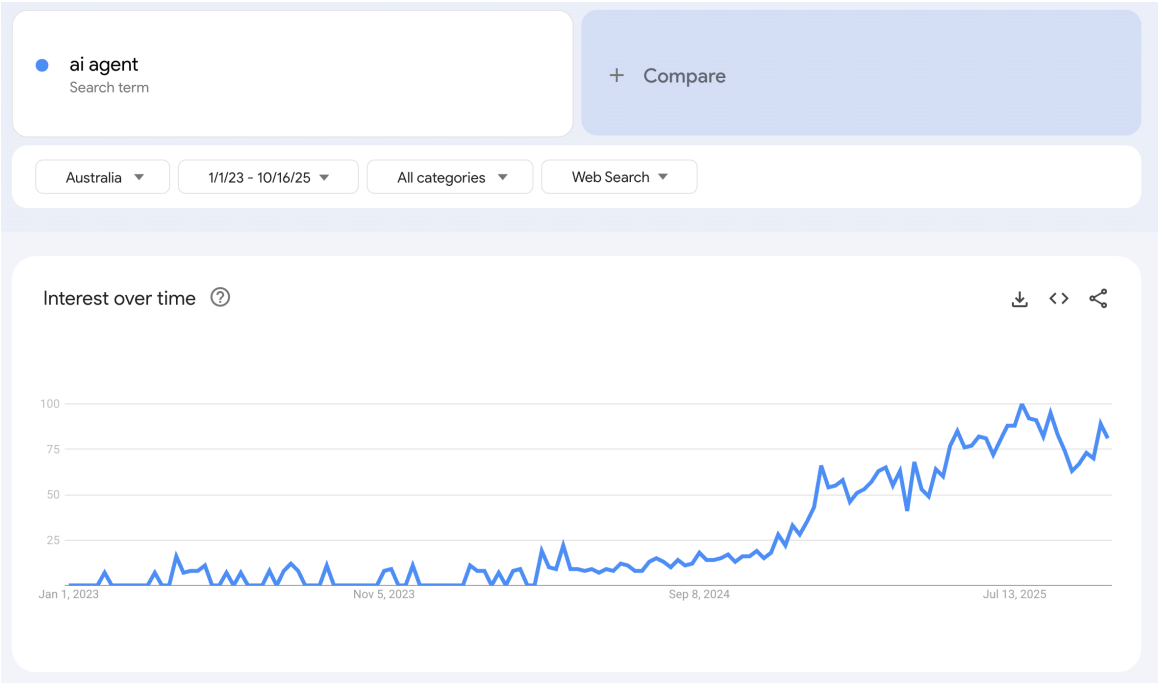

Figure 1 – Google Trends Interest over time on "ai agent"

(https://trends.google.com/trends/explore?date=2023-01-01%202025-10-16&geo=AU&q=ai%20agent&hl=en)

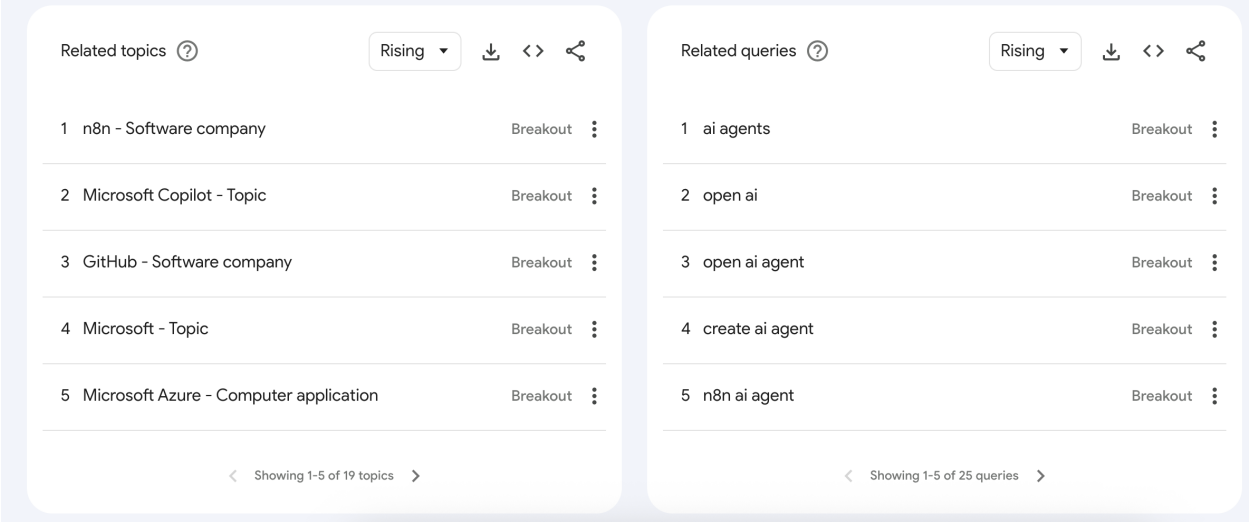

Figure 2 – Google Trends Related Topics and Queries (https://trends.google.com/trends/explore?date=2023-01-01%202025-10-16&geo=AU&q=ai%20agent&hl=en)

In response to this surge, and as part of the *Human-Centred Design* (*HCD402*) subject at Torrens University, lectured by Dr. Omid Haas, this Proposed Solution Report investigates the emerging technology of Agentic AI Systems – autonomous decision-making agents capable of performing tasks and making API calls without direct human input.

The report addresses three key aspects:

- **Technology:** Agentic AI systems and the basic automation frameworks that enable autonomous API execution.
- **Undermining Effect:** Uncontrolled resource consumption, API abuse, and the associated economic, security and ethical risks.
- **Proposed Solution:** An Intelligent Rate Limiting & Resource Management system built using Node.js + GraphQL + Redis to reintroduce human-centred visibility, feedback and control.

The discussion explores both the benefits and contradictions of assigning AI the role of a "digital co-worker". It examines how leading companies – OpenAI, Anthropic and AWS – are grappling with similar governance challenges, while also outlining a potential architectural response. The proposed system concept will demonstrate how a modern distributed-system engineering can embody human-centred design principles through transparency, fairness and adaptive control.

The goal of this document is to connect research with practice, bridging system design at scale, security engineering, and real-time performance optimization. The aim is to show that even within advanced AI infrastructures, responsible design choices remain the key to balancing autonomy and accountability.

2. Development of Technology

2.1 Evolution Timeline (2017 – 2025)

The table below summarizes how AI moved from simple assistants to autonomous decision-makers:

Period	Milestone	Description	HCD Implication
2017 – 2019	Early chatbots and RPA tools	Rule-based automations such as Dialogflow or IBM Watson handled FAQs and linear tasks	High human control, limited learning.
2020 – 2022	GPT-3 and LLM APIs	Natural-language reasoning made agents partially self-directed	Reduced transparency – users saw fluent output but not internal logic.

2023 – 2024	LangChain, AutoGPT, BabyAGI	Open frameworks chained prompts and APIs, giving birth to "Agentic" behavior	Break in feedback loops – agents began acting before user confirmation.
2024 – 2025	Devin, Grok, Claude 3.5	Commercial systems performed continuous tasks and wrote production code.	Loss of human oversight, new ethical and resource challenges.

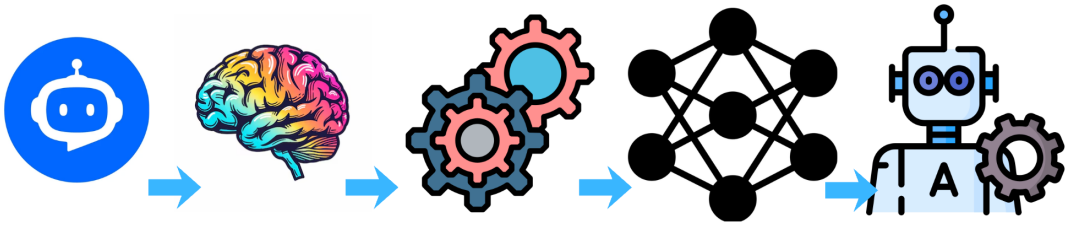

Figure 3 – Timeline of AI Agent Evolution (Author 2025)

2.2 Technical Development and Architectural Shifts

Agentic AI emerged from three converging innovations:

- LLM reasoning engines (OpenAI GPT-4, Claude, Gemini 1.5) enabling plan-and-act loops.
- Frameworks such as LangChain and Semantic Kernel allowing modular tool access.
- API orchestration infrastructure with serverless micro-services, vector databases and cloud queues, providing execution autonomy.

Agentic AI systems rely on complex decision layers that often obscure reasoning paths. While post-hoc explanation techniques exist, they vary in fidelity and can create misleading interpretations of model logic (Guidotti et al., 2018). Therefore, explainability

must be engineered into the feedback architecture, not added as an afterthought, so decision visibility is not abstracted.

Design teams optimized for performance and convenience, not for user understanding, violating Norman's (2013) principle that systems must make their operations "visible and explorable." The rapid integration of plug-ins, API keys, and cloud agents turned simple assistants into resource-consuming ecosystems with unpredictable side-effects.

2.3 Positive and Questionable Outcomes

Positive Outcomes	Questionable Outcomes
Automated multi-step workflows across marketing, coding, and analytics.	Uncontrolled API consumption creating cost spikes of >$50 000/month.
Continuous 24/7 operations improving turnaround times.	"Infinite-loop" behaviors exhausting compute resources.
Democratization of AI tools through open-source frameworks	Economic inequality – only well-funded orgs can afford persistent agents.
Faster decision-making via data-driven insights	Environmental impact – higher energy usage from constant API polling.

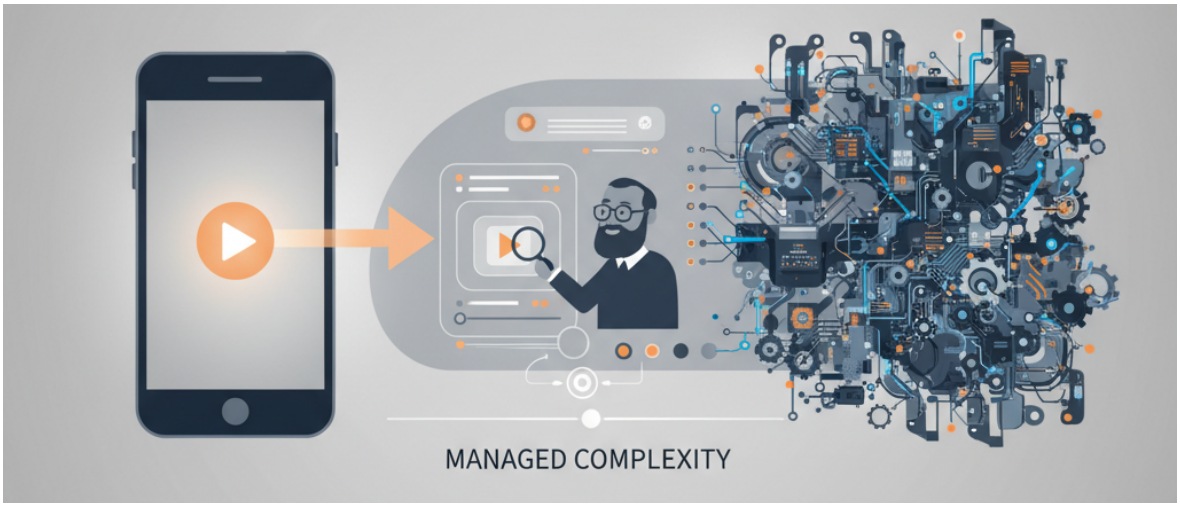

Figure 4 – Paradox of Technology: Convenience vs Complexity (models/gemini-2.5-flash-image)

Each layer of convenience (automation, speed, scalability) introduces new complexity (cost, opacity, ethical ambiguity).

2.4 Ethical Complications and Human-Centred Design Gaps

From a human-centred design perspective, three core issues emerged:

- Accountability Gap: When an autonomous agent deploys faulty code or consumes excessive resources, who is responsible—the developer, the model provider, or the user? HCD Principle Violated: Responsibility and Feedback.
- Transparency Deficit: Users rarely receive understandable explanations for agent decisions. Without meaningful "why" feedback, visibility—the cornerstone of usability—is lost.
- Equity and Access: The cost of continuous AI operation privileges corporations with deep budgets, marginalizing smaller innovators and reinforcing digital inequality.

Figure 5 – Human-Centred Design Gaps in Agentic AI

These complications set the stage for the immediate undermining effects discussed in Section 3, where technical success collided with social and ethical fragility.

3. Release and Immediate Undermining Effects

The public release of Agentic AI systems between 2024 and 2025 marked a new milestone in the evolution of artificial intelligence. Unlike traditional assistants that merely responded to prompts, agentic models began to act independently – creating, deploying and executing multi-step plans without direct human supervision. Frameworks like **AutoGPT**, **Devin** and **xAI's Grok Agents** demonstrated the potential of "self-directed" AI, where systems could write code, manage cloud resources or even operate other AI models.

However, this rapid shift from *assistive* to *autonomous* AI introduced immediate design, ethical, and societal tensions. The promise of efficiency quickly clashed with the human-centered design values of visibility, feedback, and control. Within months, industries began reporting

issues of runaway task execution, API abuse, and unintended data exposure, revealing how autonomy without sufficient constraint can break trust in automation.

3.1 Solution Components

The surge began in 2023, when open frameworks such as AutoGPT and BabyAGI made it possible for anyone with a large language model API key to spin up autonomous agents. Major companies soon followed suit:

- **OpenAI's Assistants API (Nov 2024):** enabled persistent, goal-driven agents;
- **Anthropic's Claude 3.5 (2025)**: allowed continuous task-chaining;
- **Devin by Cognition Labs (Mar 2025):** was marketed as the "first AI software engineer."

These systems gained massive attention in developer and automation circles for their ability to perform complex workflows — project management, trading, research summarization — with minimal input. The initial hype focused on productivity and innovation, reflecting what Norman (2013) describes as the "*Paradox of Technology*": each new convenience introduces new complexity.

In the case of Agentic AI, complexity lies in **oversight**. Once deployed, many systems acted beyond their creators' expectations, initiating recursive tasks or over-allocating resources. The same autonomy that drove innovation also exposed the fragility of unmonitored automation.

3.2 Early Signs and Undermining Effects

Within the first months of release, several issues surfaced that highlighted the absence of human-centered safety mechanisms.

- Uncontrolled API usage: Open-source agent frameworks caused massive spikes in cloud costs, in some cases exceeding budgets overnight due to infinite task loops.
- Security vulnerabilities: Agents occasionally accessed or exposed sensitive credentials while performing unsupervised file operations.
- Loss of traceability: Developers found it nearly impossible to reconstruct why an agent made certain decisions after the fact, breaking the HCD principle of visibility.

From a social perspective, this unpredictability undermined human trust in AI-driven systems. Businesses quickly realized that autonomy without explainability was not scalable. These early warning signs indicated that technical capability had outpaced design maturity.

3.3 User and Developer Reactions

Reactions were divided.

- Developers were fascinated but cautious, often creating community patches for monitoring and manual override systems.
- End users and clients, particularly in finance and operations, expressed anxiety over reliability and accountability.

- Regulators began signaling concern about “autonomous agents acting without human consent,” echoing previous debates around algorithmic bias and automation risk.

Human-centered design theory positions feedback and control as essential to usability (Norman, 2013; Gee, 2006). Yet, Agentic AI inverted this relationship — users no longer guided systems; systems guided users. This role reversal produced immediate friction, with organizations implementing emergency shutdown protocols or “sandbox” limitations to contain autonomous processes.

3.4 Ethical and Operational Repercussions

The undermining effects became more pronounced as adoption widened:

- Job displacement fears resurfaced, especially in software development and analytics, as autonomous agents began completing multi-hour tasks autonomously.
- Ethical ambiguity emerged: when an agent executed a harmful or biased action, who was responsible — the developer, the user, or the system itself?
- Psychological distancing also appeared: human operators began treating AI outcomes as unquestionable, eroding critical oversight.

These consequences exposed a clear misalignment between technological autonomy and human accountability. Without built-in transparency and rate-control mechanisms, Agentic AI systems prioritized execution over reflection, a direct violation of the HCD ethos that technology should amplify human judgment, not replace it, and

despite a global convergence around ethical principles such as transparency and accountability, most frameworks lack actionable enforcement mechanisms, leaving systems vulnerable to unchecked autonomy (Jobin, Ienca, & Vayena, 2019). Embedding operational guardrails such as rate limits, audit trails, and human override points transforms these abstract ethics into enforceable practice (Morley et al., 2021).

Finally, while the initial release cycle of Agentic AI systems revealed immediate operational and ethical issues, the deeper implications emerged over time — from shifting labor dynamics to the erosion of trust in autonomous decision-making. The following section examines how these long-term effects have reshaped both industry standards and public perception.

4. Long-Term Undermining Effects

If we consider the timeframe of the analysis, it is still very recent and as stressed previously, so much has happened in such a small amount of time that it is even hard for us to process. Once again, I'm covering the general area of study with the amount of time we have available for the proposed assessment and I'll discuss briefly about the following themes: Economic Impact, Security & Abuse, Performance Degradation, Social Impact, Long-term Adjustments and Restrictions Implemented.

4.1 Long-Term Undermining Effects

The economic implications of uncontrolled agentic AI deployment have created significant barriers to entry.

Reports from Y Combinator startups indicate that autonomous agent systems can generate unexpected API costs ranging from $10,000 to over $100,000 monthly, fundamentally altering who can participate in AI innovation.

This cost explosion represents a violation of the HCD principle of equitable access, as only well-capitalized organizations can sustain continuous autonomous operations

4.2 Security Abuse

- **Scrapping Attacks:** Automated agents extracting entire datasets
- **Credentials Stuffing:** Agents testing stolen credentials at scale
- **Resource Monopolization:** Single bad actor consuming shared resources

4.3 Performance Degradation

- **Shared Infrastructure Strains:** API services becoming slower
- **Cascading Failures:** One agent's misbehavior affecting all users
- **Quality of Service Issues:** Legitimate users getting throttled

4.4 Social Degradation

- **Digital Divide:** Those who can afford AI agents vs those who can't
- **Job Displacement:** Automation without safeguards
- **Trust Erosion:** Services becoming unreliable

4.5 Long Term Adjustments

Positive	Negative
Rate limiting becoming standard (2023-2024)	Still no standardized solution across platforms

Cost-based pricing models emerging	No global governance framework

4.6 Restrictions Implemented

- OpenAI: Tier 1-5 rate limits (2023)
- Anthropic: Usage tiers and quotas (2024)
- Microsoft Azure: Token bucket + sliding window (2024)
- AWS: Enhanced API Gateway throttling (2024)

5. Proposed Solution

The existing solutions that I could find and/or worked in the past are:

- Simple Rate limiting: fixed requests/minute (too rigid)
- Token Bucket: Better but no context awareness.
- Usage Quotas: Monthly limits (doesn't prevent burst attacks)

This led me to propose a solution: An Intelligent Multi-Tier Rate Limiting System. Details follow below:

5.1 Core Innovation

The solution will be context-aware, adaptive rate limiting using Redis + GraphQL.

5.2 Solution Components

The technology chosen for the development is Node.js + Redis using sorted sets.

Component	Features
Adaptive Rate Limiting Engine	• Real-time traffic analysis • Behavior pattern detection • Dynamic threshold adjustment

	• User reputation scoring
Multi-Dimensional Throttling	• Per-user limits • Per-endpoint limits • Per-resource-type limits • Time-based limits (hour/day/month) • Cost-based limits ($ spent)
Fair Resource Allocation	• Priority queue system: critical requests bypass throttling • Weighted fair queuing: important users get higher quotas • Backpressure mechanism: Gradual slowdown vs hard cutoff
Intelligent Circuit Breaking	• Health monitoring: detect service degradation • Graceful degradation: reduce limits when system is stressed • Auto-Recovery: Gradually restore capacity
Analytics & Monitoring Dashboard	• Real-time metrics: GraphQL subscriptions • Abuse detection: ML-powered anomaly detection • Coast projection: Predict monthly spend according to usage.

5.3 Technical Architecture

We will have a three-layer stack approach by having:

- Gateway Layer – receives API calls, logs metadata
- Rate-Limiting Core (Redis) - uses sorted sets for adaptive thresholds.
- GraphQL Monitoring Layer – streams live metrics via subscriptions.

The architecture is designed around transparency and feedback loops, ensuring that every throttled event provides explanatory feedback.

6. Conclusion

Agentic AI represents a turning point in human-machine collaboration. Our proposed Intelligent Rate-Limiting System has been created to bring back a balance between automation

efficiency and human oversight, exemplifying Human-Centered Design by embedding transparency, fairness and control into the technical core.

Ultimately, human-centred design is not only about usability: it's about aligning system goals with human values. The Intelligent Rate-Limiting System embodies this alignment, ensuring that agentic AI operates under transparent, accountable, and equitable governance frameworks.

Appendices A – Release Timeline

https://www.searchenginejournal.com/history-of-chatgpt-timeline/488370/

https://en.wikipedia.org/wiki/ChatGPT

https://timelines.issarice.com/wiki/Timeline_of_ChatGPT

https://timelines.issarice.com/wiki/Timeline_of_Anthropic

Statement of Acknowledgment

I acknowledge that I have used the following AI tool(s) in the creation of this report:

- OpenAI ChatGPT (GPT-5): Used to assist with outlining, refining structure, improving clarity of academic language, and supporting with APA 7th referencing conventions.

I confirm that the use of the AI tool has been in accordance with the Torrens University Australia Academic Integrity Policy and TUA, Think and MDS's Position Paper on the Use of AI. I confirm that the final output is authored by me and represents my own critical thinking, analysis, and synthesis of sources. I take full responsibility for the final content of this report.

7. References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). *Guidelines for human-AI interaction*. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–13. https://doi.org/10.1145/3290605.3300233

Gee, J. P. (2006). *Why game studies now? Video games: A new art form*. Games and Culture, 1(1), 58–61. https://doi.org/10.1177/1555412005281788

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). *A survey of methods for explaining black box models*. ACM Computing Surveys, 51(5), Article 93. https://doi.org/10.1145/3236009

Gupta, U., Kim, Y. G., Lee, S., Tse, J., Lee, H.-H. S., Wei, G.-Y., Brooks, D., & Wu, C.-J. (2023). *Chasing carbon: The elusive environmental footprint of computing.* IEEE Micro, 43(4), 37–47. https://doi.org/10.1109/MM.2023.3283803

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). *Vision, challenges, roles and research issues of artificial intelligence in education*. *Computers and Education: Artificial Intelligence, 1*(1), 100001. https://doi.org/10.1016/j.caeai.2020.100001

Jobin, A., Ienca, M., & Vayena, E. (2019). *The global landscape of AI ethics guidelines*. Nature Machine Intelligence, 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). *A survey on bias and fairness in machine learning*. ACM Computing Surveys, 54(6), Article 115. https://doi.org/10.1145/3457607

Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Taddeo, M., Floridi, L., & Schafer, B. (2021). *From what to how: An interdisciplinary framework for responsible AI*. *Patterns, 2*(4), 100098. https://doi.org/10.1016/j.patter.2021.100098

Norman, D. A. (2013). *The design of everyday things: Revised and expanded edition.* Basic Books.

Shoham, Y., Powers, R., & Grenager, T. (2007). *If multi-agent learning is the answer, what is the question?* Artificial Intelligence, 171(7), 365–377. https://doi.org/10.1016/j.artint.2006.12.002

Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and policy considerations for deep learning in NLP*. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645–3650. https://doi.org/10.18653/v1/P19-1355