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

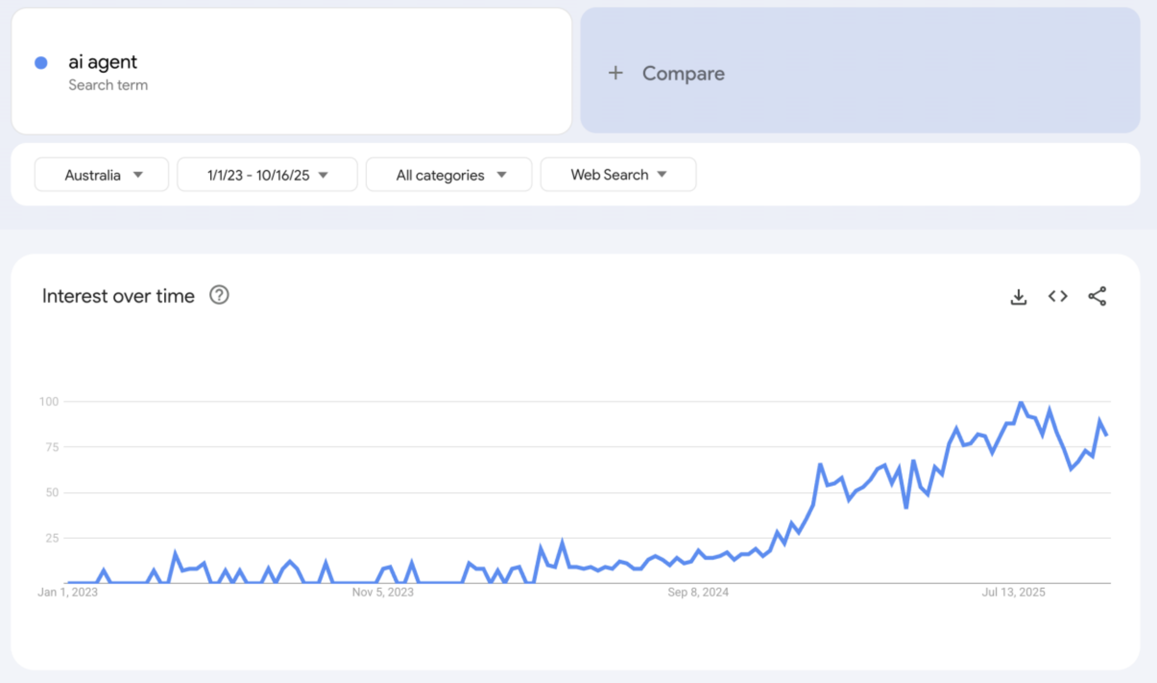
Lecturer: Dr. Omid Haas

Date: Nov 2025

Table of Contents

1. Introduction

It is impossible not to hear about AI agents nowadays, we read about them on the news, saying Artificial Intelligence is replacing our jobs, we read about people talking about them on LinkedIn (‘Comment ‘AGENT’ on this post to receive the step by step’). The picture below demonstrates the increase of search on that term on the past 2 years (Jan 2023 – Oct 2025):



ai agent
Search term

Australia | 1/1/23 - 10/16/25 | All categories | Web Search

Fig 1 – Google Trends (https://trends.google.com/trends/explore?date=2023-01-01%202025-10-16&geo=AU&q=ai%20agent&hl=en)

Based on that crazy demand, for the proposed assignment of Human-Centred Design subject at Torrens University, being lectured by Dr. Omid Haas, I have decided to write this Proposed Solution Report with the following characteristics:

- **Technology:** Agentic AI systems (autonomous AI agents making API calls)

- **Undermining Effect:** Uncontrolled resource consumption, API abuse, economic/security risks.
- **Proposed Solution:** Intelligent Rate Limiting & Resource Management system using Node.js + GraphQL + Redis.

I intend to bring the reader deep with me on the benefits and contradictions of having AI as our workers and discuss the fact that companies are losing thousands of dollars (not to mention reputation) on uncontrolled AI agent API calls. We'll discuss about OpenAI, Anthropic, AWS solutions implemented and we'll also dip our toes on the waters of a possible Node.JS + Apollo Server + Redis solution development that will challenge and deepen our knowledge with a cutting edge API system built using advanced Redis patterns, Graph QL subscriptions for real-time monitoring, Entreprise grade middleware architecture and distributed systems design.

This will demonstrate system design at scale, security engineering, performance optimization, real-time systems and research to production link. I hope you enjoy the trip below. It has been interesting to write about this.

2. Development of Technology

Timeline 2017-2024

- 17-19: Early AI Assistants (chatbots, simple automation)
- 20-22: GPT-3 enables more autonomous behavior
- 23-24: Full agentic systems (AutoGPT, LangChain agents, etc.)

Main Effects:

Positive	**Questionable**
Automation of complex workflows	Uncontrolled API consumption
Enhanced productivity and decision-making	Resource exhaustion attacks
24/7 autonomous operations	Economic inequality (who can afford unlimited API access)

Development outcomes

- Inspired microservices architecture evolution
- Led to serverless computing adoption
- Drove need for better API management

Ethical Complications

- Who's responsible when an agent causes harm?
- How to prevent malicious agent deployment?
- Fair resource allocation among users

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. Release Timeline

The surge began in early 2024, 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 Warning 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.

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. Economic Impact

- **Cost Explosion:** Startups facing $10K-$100K monthly API bills
- **Barrier to Entry:** Only well-funded companies can afford agentic systems
- **Market Consolidation:** Large players dominate due to API access advantages

4.2. Security and 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 Impact

- **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)

Blabla blab la bla.

5. Proposed Solution

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

1. Simple Rate limiting: fixed requests/minute (too rigid)
2. Token Bucket: Better but no context awareness.
3. 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 solution will be context-aware, adaptive rate limiting using Redis + GraphQL and 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:

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

The architecture was 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 restores balance between automation efficiency and human oversight, exemplifying Human-Centered Design by embedding transparency, fairness and control into the technical core.

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

Alkhnbashi, O. S., Mohammad, R., & Hammoudeh, M. (2024). *Aspect-based sentiment analysis of patient feedback using large language models*. *Big Data and Cognitive Computing*, *8*(12), 167. https://doi.org/10.3390/bdcc8120167

Angelis, J. N., Murthy, R. S., Beaulieu, T., & Miller, J. C. (2024). *Better angry than afraid: The case of post data breach emotions on customer engagement*. *IEEE Transactions on Engineering Management*, *71*, 2593–2605. https://doi.org/10.1109/TEM.2022.3189599

Chen, E. (2023). *Growth product manager's handbook.* O'Reilly Media.

Dawes, J. G. (2024). *The net promoter score: What should managers know? International Journal of Market Research*, *66*(2–3), 182–198. https://doi.org/10.1177/14707853231195003

Godovykh, M., & Pizam, A. (2023). *Measuring patient experience in healthcare*. *International Journal of Hospitality Management*, *112,* 103405. https://doi.org/10.1016/j.ijhm.2022.103405

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

Mar, J., & Armaly, P. (2023). *Mastering customer success.* O'Reilly Media.

Shankar, R., & Yip, A. (2024). *Transforming patient feedback into actionable insights through natural language processing: A knowledge discovery and action research study. JMIR Formative Research*. *Advance online publication.* https://doi.org/10.2196/69699

Xiao, Y., Li, C., Thürer, M., Liu, Y., & Qu, T. (2022). *Towards lean automation: Fine-grained sentiment analysis for customer value identification. Computers & Industrial Engineering*, *169*, 108186. https://doi.org/10.1016/j.cie.2022.108186