

INTELLIGENT RATE LIMITTING SYSTEM

Restoring Human Agency in Autonomous AI Systems

SYSTEM SOLUTION | ASSESSMENT 3
HUMAN-CENTRED DESIGN
PROF DR. OMID HAAS

Group #2

Luis G. B. A. Faria – ID A00187785 [Technical Architecture]

Julio Ibanez Bertrand – ID A00119197 [Ethical Frameworks]

Tamarra Berryman – ID A00205009 [HCD Integration & Social Impact]



December, 2025

THE RISE OF AGENTIC AI → A NEW RISK FRONTIER

- Autonomous AI agents (AutoGPT, Devin, Grok) now execute multi-step tasks without human supervision.
- API-driven autonomy creates uncontrolled resource consumption, cost explosions, and system instability.
- Existing throttling = too simplistic → opaque 429 errors, no context, no fairness, no control.
- Human-Centred Design gap: users lack visibility, predictability, agency, and trust.

WHAT GOES WRONG TODAY?

Environmental Impact

Human-Centered

- Continuous agent workloads create energy spikes.
- No carbon-aware routing → unnecessary CO₂ emissions.



Performance Instability

Wellbeing & Addiction

- API overloads → cascading failures.
- Tool-use loops cause “runaway agents”.

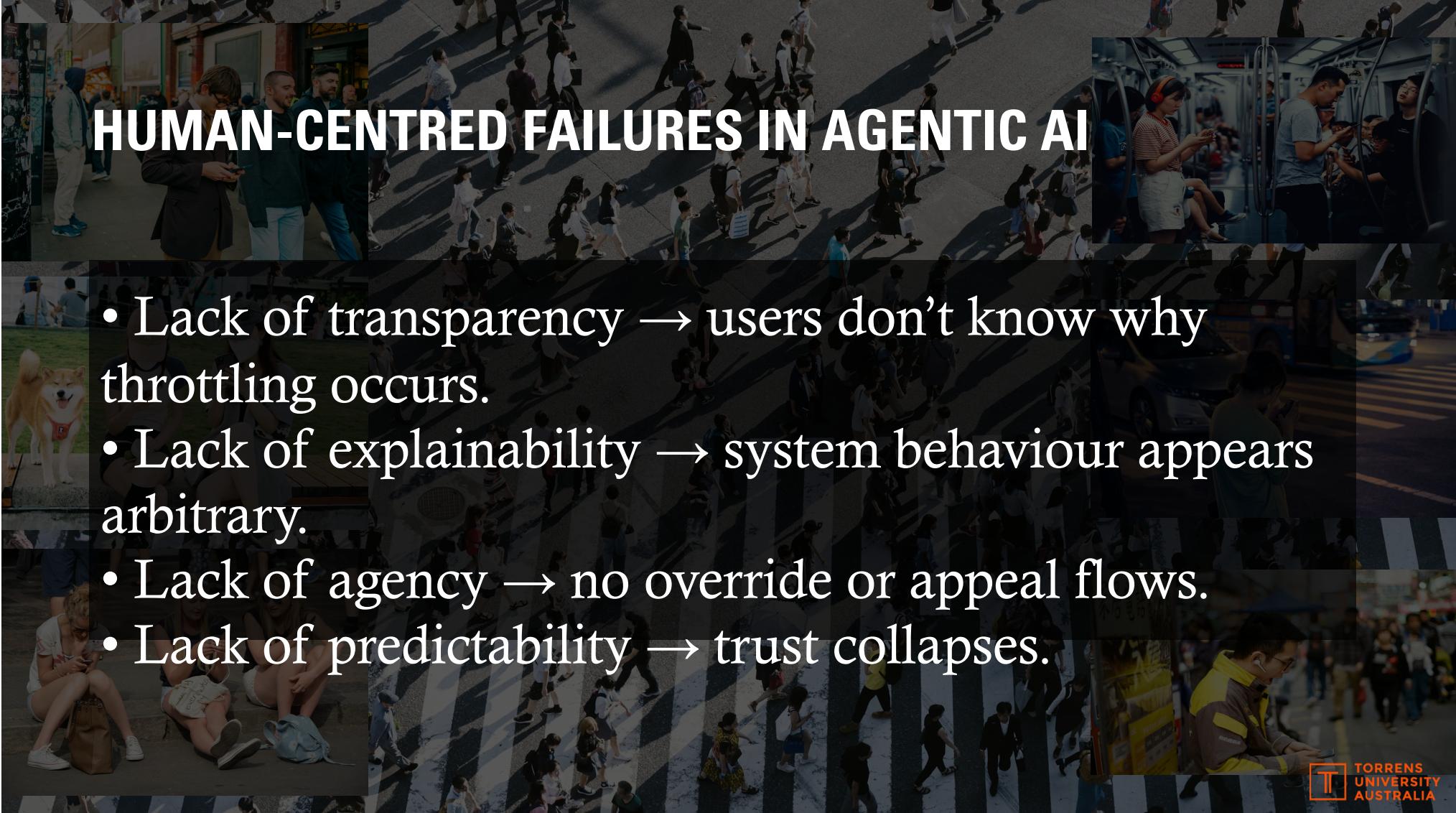


Fairness & User Impact

Governance & Data

- Opaque throttling → developer confusion.
- Perceived unfairness → loss of trust in AI systems.





HUMAN-CENTRED FAILURES IN AGENTIC AI

- Lack of transparency → users don't know why throttling occurs.
- Lack of explainability → system behaviour appears arbitrary.
- Lack of agency → no override or appeal flows.
- Lack of predictability → trust collapses.

THE INTELLIGENT RATE LIMITING SYSTEM

IRL is a multi-tier, adaptive throttling system that gives developers visibility, context, fairness and agency while reducing environmental and financial costs.

Dynamic Rate Limiting

Carbon-Aware Throttling

Weighted-Fair Queuing

Explainable Feedback Layer

ADAPTIVE QUOTAS WITH CONTEXTUAL FEEDBACK

CORE COMPONENT #1

- Real-time monitoring of tokens, cost, frequency
- No more opaque 429 errors
- Contrastive explanations (e.g., “why blocked + how to succeed”)
- Human-in-the-loop override

CARBON-AWARE THROTTLING

CORE COMPONENT #2

- Integrates grid carbon intensity API
- Low-renewable periods → deprioritize non-urgent tasks
- Shows carbon + dollar impact side by side

CARBON-AWARE THROTTLING

CORE COMPONENT #3

- Integrates grid carbon intensity API
- Low-renewable periods → deprioritize non-urgent tasks
- Shows carbon + dollar impact side by side

DESIGNING FOR VISIBILITY, PREDICTABILITY & CONTROL

CORE COMPONENT #4

- Real-time quotas & usage
- Estimated reset time
- Override button
- Carbon & financial impact
- Alert center

PREDICTING EFFECTIVENESS USING HCI & XAI RESEARCH

- Transparency: follows XAI principles (Guidotti et al., 2018; Miller, 2019).
 - Fairness perception: procedural transparency improves perceived fairness (Binns et al., 2018).
 - Trust: contrastive explanations boost trust in autonomous systems.
 - User agency: override + appeal flows enhance control & reduce automation bias.
-

LIMITATIONS

- Requires accurate carbon intensity data
- Needs broad user testing across cultures & technical skill levels
- Policy alignment required for enterprise adoption

FUTURE WORK

- Multi-modal agent oversight (vision + text)
- Predictive anomaly detection using RAG
- Usability studies with cross-regional developers
(aligned to Triandis)

CONCLUSION

IRL is a human-centered, transparent, sustainable and fair solution to one of the biggest problems introduced by Agentic AI: uncontrolled autonomy.

It restores trust, reduces cost, minimizes CO₂ impact, and empowers human oversight.

STATEMENT OF ACKNOWLEDGEMENT

We acknowledge that we have used OpenAI's ChatGPT (GPT-5) to assist in the planning, outlining, and refinement of my presentation for HCD402 – Assessment 3. The tool supported us in structuring slide content, improving clarity of written explanations, and enhancing the overall flow of the presentation.

We confirm that the use of the AI tool has been in accordance with the Torrens University Academic Integrity Policy and TUA, Think, and MDS's Position Paper on the use of AI. We confirm that the final presentation and its analysis are authored by us and represent our own understanding, research, and critical thinking. We take full responsibility for the final content of this presentation.

REFERENCES

- Alevizos, V., Gerolimos, N., Lelikou, E. A., Hompis, G., Priniotakis, G., & Papakostas, G. A. (2025). *Sustainable swarm Intelligence: Assessing carbon-aware optimization in high-performance AI systems*. Technologies, 13(10), 477. <https://doi.org/10.3390/technologies13100477>
- 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>
- Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). *'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions*. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1–14. <https://doi.org/10.1145/3173574.3173951>
- Green Software Foundation. (2024). *Carbon Aware SDK documentation*. <https://carbon-aware-sdk.greensoftware.foundation/>
- 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.caai.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>
- Miller, T. (2019). *Explanation in artificial intelligence: Insights from the social sciences*. Artificial Intelligence, 267, 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- 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>
- 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>
- Triandis, H. C. (2018). *Individualism and collectivism* (Reissued ed.). Routledge.
- Wiesner, P., Behnke, I., Scheinert, D., Gontarska, K., & Thamsen, L. (2023). *Let's wait awhile: How temporal workload shifting can reduce carbon emissions in the cloud*. Proceedings of the 22nd International Middleware Conference, 260–272. <https://doi.org/10.1145/3590140.3629116>
- Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., Yin, Z., Dou, S., Weng, R., Cheng, W., Zhang, Q., Qin, W., Zheng, Y., Qiu, X., Huang, X., & Gui, T. (2024). *The rise and potential of large language model based agents: A survey*. arXiv preprint arXiv:2309.07864. <https://doi.org/10.48550/arXiv.2309.07864>

SYSTEM SOLUTION
ASSESSMENT 3

HUMAN-CENTRED DESIGN
PROF DR. OMID HAAS

Luis G. B. A. Faria – ID A00187785
Julio Ibanez Bertrand – ID A00119197
Tamarra Berryman – ID A00205009



Thank You

December, 2025

