AGI NEEDS HUNGER: METABOLIC GROUNDING AS A CONSTRAINT ON ARTIFICIAL AGENCY

Sahil Gupta
Department of Computer Science
Brown University
Providence, RI 02912, USA
sahil_gupta@brown.edu

ABSTRACT

Current AI systems exhibit impressive pattern recognition, reasoning, and planning, yet remain fundamentally *unmotivated*: they optimize externally specified objectives without intrinsic pressure to persist, adapt, or care. We argue that this limitation reflects a missing constraint shared by all biological intelligence: *metabolic scarcity*. In living systems, cognition is inseparable from energy regulation. Biological agents must continuously act to maintain viability. Building on biological computationalism, which views cognition as embodied computation realized through physical self-maintenance, we advance the position that artificial general intelligence will require an analogous internal energetic imperative ("artificial hunger") to generate real agency, self-directed learning, and consciousness. We propose a framework for *metabolically grounded computation*, outline architectural guidelines, and derive falsifiable predictions for future AI systems.

1 INTRODUCTION: INTELLIGENCE WITHOUT NEED

Deep learning systems now rival or exceed humans in domains such as vision (Radford et al., 2021), protein folding (Abramson et al., 2024), and strategic games (Silver et al., 2016). However, these systems remain *instrumental optimizers*: they pursue reward functions and loss objectives defined from the outside. They do not exhibit the core property of biological agents: *acting under internal necessity*. The human brain is a prime example: despite making up only 2% of body mass, it expends 20% of the body's energy budget (Raichle & Gusnard, 2002). For biological organisms, every unit of energy expended on cognition is energy unavailable for other survival-critical functions like immunity and tissue repair. This creates genuine resource scarcity at every level, from individual neurons competing for glucose to large-scale brain networks negotiating for metabolic resources.

Contemporary AI systems, by contrast, face no such intrinsic constraints. While LLM inference operations consume electrical power, these costs are entirely externalized to operators (e.g. OpenAI). When GPT-4 generates a response, there is no internal pressure toward efficiency, no trade-off between exploring novel approaches versus exploiting known solutions. Resource limits are imposed externally through context windows and rate limiting rather than from thermodynamic constraints.

Position: We believe the difference between intrinsic vs extrinsic grounding is fundamental to machine consciousness. Not because consciousness requires carbon-based biology specifically, but because consciousness may require a particular form of computational organization that emerges only under genuine energetic constraint. Thus, **true AGI will require an internal analogue of biological hunger: a system-level drive arising from energetic scarcity and self-maintenance.**

2 BACKGROUND: BIOLOGICAL COMPUTATIONALISM

Biological computationalism (Milinkovic & Aru, 2026) proposes that cognition is not abstract symbol manipulation but a form of computation inseparable from the biological substrate. Brains do not compute in isolation: they compute as part of an organism and organ system struggling to persist. Cognition is therefore fundamentally *metabolic*. Organisms must regulate internal variables (temperature, pH, ion concentration, etc) through specific action and complex feedback mechanisms.

This is known as *allostasis*: predictive regulation in the service of survival (Sterling, 2012). Agency emerges because the organism must continually secure the conditions of its own continuation.

This biological grounding may also connect to consciousness. Several theories formalize this claim, claiming that emotions are responses to an organism's change in homeostasis (Fokas & Logothetis, 2025), and consciousness is fundamentally driven by the emergence of homeostatic feelings like hunger, pain, desire, and thirst (Damasio & Damasio, 2022). Under this view, optimization via user-defined reward functions may be insufficient for conscious AGI. We do not claim that metabolic grounding *necessarily* produces consciousness, but we do argue that it constitutes a *prerequisite structural condition* that current architectures lack.

3 Why Modern AI Lacks Biological Grounding

Biological and AI systems differ in their relationship to energetic cost. Neural computation is inherently energy-limited. When neurons fire repeatedly, they deplete local ATP reserves and accumulate metabolic byproducts. This creates "metabolic memory" where current capacity reflects computational history. Critically, metabolic constraints create conditions for genuine valuation to emerge. Some computations become more costly, and performing them reduces future capacity. A human may choose not to stay up the night before taking a big exam to preserve mental function. This creates intrinsic pressure toward efficiency. A metabolically constrained system cannot waste resources on task-irrelevant operations because doing so threatens future necessary computations.

Current AI systems have no such intrinsic limitation. Modern systems have external constraints: context windows limit token count, rate-limiting caps throughput, and hardware budgets bound wall-clock time. However, these constraints are *episodic*, while biological metabolic constraints are *continuous*. Notably, extrinsic constraints do not feed back into the agent's internal state. When an LLM generates its 1000th token, the computational cost to the model is indistinguishable from generating its 1st token; future computations do not functionally rely on past ones. Computational resources are unlimited from the system's perspective. The algorithm itself neither monitors nor responds to its own resource consumption, operating at full capacity until it cannot operate at all.

Empirical work on intrinsically motivated reinforcement learning (RL) further supports this distinction. Chentanez et al. (2004) demonstrated that agents equipped with internal reward signals—generated from within the agent's own architecture—acquire hierarchical skills and exhibit persistent exploratory behavior that agents relying solely on external rewards do not. Critically, the source of the signal (internal vs. external) produces qualitatively different learning dynamics. More recent work by Villalobos-Arias et al. (2025) finds that intrinsic motivation methods, which are grounded in psychological concepts such as curiosity, empowerment, and novelty, produce systematic behavioral changes in RL agents that extrinsic reward shaping alone does not replicate.

3.1 Implications for Current AI Limitations

A system that does not experience the cost of its own operation, that faces no genuine internal trade-offs between competing computational demands, may lack the prerequisites for consciousness. The absence of intrinsic motivation in modern AI systems may help explain gaps in today's landscape:

- **Hallucinations.** LLMs often produce fluent but factually incorrect outputs. Philosophers and cognitive scientists have long argued that meaningful understanding is relational: symbols gain meaning through embodiment and interaction with the world (Harnad, 1990). Because LLMs optimize predictive likelihood on text rather than homeostatic viability or sensorimotor contingencies, there is no pressure for their internal states to be accountable to real-world structure. Take health-related guardrails — LLMs only refrain from providing health advice because of external rules imposed by their creators. The LLM itself has limited understanding of how wrong advice may affect the user. In nature, biological agents cannot hallucinate arbitrarily: incorrect models of the world carry metabolic cost and can threaten survival of both self and others. The absence of an embodied grounding may explain why digital systems can confidently assert false facts: truth in their world is defined by statistical co-occurrence, not survival-relevant consistency with an external environment.

- **Fragile autonomy.** Current models excel in narrowly defined tasks but struggle with open-ended goal pursuit outside of training distributions. For example, embodied AI agents trained in simulated environments often fail to generalize to new worlds or long-horizon tasks without retraining. This fragility may stem from the fact that such agents lack a notion of self-continuity; there is no internal cost of "death" or depletion that constrains their actions. In contrast, biological organisms evolved behaviors that balance exploitation and exploration across lifetimes because failure to secure resources entails existential risk.
- **Lack of continual, self-directed learning.** Biological organisms exhibit lifelong learning that incorporates new information while preserving essential competencies. In contrast, many AI systems suffer from catastrophic forgetting when trained sequentially on new tasks. Continual learning frameworks mitigate this through architectural tricks (e.g., elastic weight consolidation), but they do not imbue systems with a principled reason to preserve knowledge. Biological neural systems, by contrast, organize memory around survival-critical domains, preserving information that historically affects fitness and guides evolutionary decisions. The absence of an internal survival budget in artificial agents means learning lacks the regulatory pressure that anchors knowledge retention in organisms.

Across these examples, the common thread is that biological systems embed cognition in a life-preserving economy of needs. AI systems, however capable in surface tasks, lack the structural pressures that make biological cognition robust, persistent, and meaningfully grounded.

4 Toward Metabolically Constrained AGI

If metabolic constraints are indeed necessary for—or at minimum deeply conducive to—the emergence of agency and potentially consciousness, the engineering challenge shifts fundamentally. Rather than asking what algorithms should be implemented, we must ask what kind of physical systems can instantiate computation where energetic cost is intrinsic rather than external. To answer this question, we first formally define artificial hunger as follows:

Definition (Artificial Hunger / Metabolic Constraint).
Artificial hunger is the principle that an artificial agent possesses an **finite viability budget**—such as energy, entropy, or stability—that directly constrains its computation.

Formally, an agent is *metabolically grounded* if its policy $\pi(a \mid s)$ is optimized not only for external task reward, but also for maintaining an internal viability condition:

$$\pi^* = \arg\max_{\pi} \ \mathbb{E}\Big[\sum_t r_{\text{task}}(t) \ - \ \lambda\, C_{\text{viability}}(t)\Big],$$

where $C_{\text{viability}}$ penalizes trajectories that deplete internal resources below a survival threshold. Under artificial hunger, cognition is inseparable from self-maintenance: agents must allocate computation, explore, and plan in ways that preserve continued operation.

The viability cost $C_{\text{viability}}(t)$ must be computed from some internal state variable $e_t \in [0, E_{\max}]$ representing available energy. In a software implementation, e_t could be a scalar maintained in the agent's recurrent hidden state, updated each step by a consumption function $\delta(a_t, s_t)$ that subtracts energy proportional to computational complexity (e.g., attention operations, memory reads) and adds energy via "recovery" actions. The policy then receives e_t as part of its observation. Under gradient-based optimization, the penalty term $\lambda C_{\text{viability}}(t)$ backpropagates gradients that discourage trajectories that deplete e_t below a survival threshold.

Temporal dynamics are equally important. Biological systems require recovery cycles such as sleep; analogously, metabolically grounded agents may need enforced rest where computation is limited but internal stability is restored. Power-gating mechanisms (Xue & Huang, 2025) could render computational modules temporarily unavailable when reserves are depleted, forcing agents to learn strategies of prioritization and selective attention. Gupta et al. (2025) proposed periods of microsleep and longer nightly "offline" sessions where AI systems adaptively prune unnecessary weights, mimicking the brain's decay of underutilized synapses over time. While no empirical results were presented in that study, such proposals are a step in the right direction and should be explored further.

4.1 Empirical Predictions and Research Directions

The metabolic constraint hypothesis generates several empirically testable predictions that we hope can guide future research directions for achieving metabolically grounded and conscious AGI:

Prediction 1: Characteristic degradation patterns. Systems with genuine metabolic constraints should exhibit gradual reduction in processing speed, increased error rates in complex computations, and preferential preservation of critical functions at the expense of auxiliary processing as energy depletes. These patterns should mirror biological cognitive fatigue rather than the abrupt failures characteristic of conventional computational systems encountering resource limits.

Prediction 2: Spontaneous rest behavior. Without explicit programming for sleep, metabolically constrained systems should spontaneously develop temporal patterns of reduced activity for energy recovery. These rest periods should correlate with energy depletion rather than occurring on fixed schedules, and their duration should relate to the degree of prior energy expenditure. During rest, systems should show preferential consolidation of recently acquired information, analogous to memory consolidation during biological sleep.

Prediction 3: Energy-dependent risk modulation. Decision-making under uncertainty should exhibit systematic variation with energetic state. As available energy decreases, systems should demonstrate increasing risk aversion even in tasks where risk preferences were not explicitly shaped through training. This risk aversion should be context-sensitive, with systems accepting greater risk when potential rewards include energy acquisition opportunities.

Prediction 4: Novel learning dynamics. Learning trajectories for metabolically constrained systems should differ from conventional RL. We predict the emergence of more rapid acquisition of energy-efficient strategies and new methods for energy management that generalize across tasks.

Prediction 5: Emergent homeostatic behaviors. Systems should develop behaviors that maintain energy within viable ranges without these behaviors being specified in training objectives. This should include seeking energy-acquisition opportunities when reserves are low, avoiding unnecessary computation when reserves are not abundant.

These predictions could initially be tested using software simulations where metabolic constraints are implemented in agent-based models before attempting more challenging hardware implementations. Comparative studies between metabolically constrained and unconstrained agents performing identical tasks would reveal whether predicted behavioral and learning differences emerge.

4.2 Governance and Safety Implications

Metabolically grounded AI also introduces new governance-related dimensions. Systems endowed with intrinsic self-maintenance objectives may exhibit qualitatively different autonomy profiles, including persistence, self-protective behavior, and goal stability that extend beyond externally specified task optimization. Such properties introduce novel alignment and oversight challenges: agents driven by internal viability constraints could resist interruption, adapt strategically to preserve continuity, or develop emergent motivations not captured by standard reward-based evaluations.

Accordingly, research on artificial hunger should proceed in parallel with governance frameworks that (a) monitor autonomy thresholds (i.e., transitions from task pursuit to self-preservation), (b) ensure reliable containment and shutdown despite viability drives, and (c) establish evaluation standards and audits for survival-like behavioral signatures. We expect this field to mature as more research on the predictions in Section 4.2 are addressed by empirical studies and evaluations.

5 Conclusion

We argue that the path to AGI cannot be understood solely as scaling digital optimization. Biological intelligence is inseparable from metabolic scarcity: organisms think because they must. Artificial systems remain powerful but fundamentally unmotivated. Introducing artificial hunger—metabolic grounding through internal viability constraints—offers a principled framework for building more autonomous, adaptive, and biologically plausible intelligence.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016): 493–500, June 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-07487-w. URL https://www.nature.com/articles/s41586-024-07487-w.

Nuttapong Chentanez, Andrew Barto, and Satinder Singh. Intrinsically motivated reinforcement learning. In L. Saul, Y. Weiss, and L. Bottou (eds.), *Advances in Neural Information Processing Systems*, volume 17. MIT Press, 2004. URL https://proceedings.neurips.cc/paper_files/paper/2004/file/4be5a36cbaca8ab9d2066debfe4e65c1-Paper.pdf.

Antonio Damasio and Hanna Damasio. Homeostatic feelings and the biology of consciousness. *Brain*, 145(7):2231–2235, July 2022. ISSN 0006-8950, 1460-2156. doi: 10.1093/brain/awac194. URL https://academic.oup.com/brain/article/145/7/2231/6594735.

Athanassios S. Fokas and Nikos K. Logothetis. Conscious and unconscious processes in vision and homeostasis. *Frontiers in Behavioral Neuroscience*, 19:1516127, February 2025. ISSN 1662-5153. doi: 10.3389/fnbeh.2025.1516127. URL https://www.frontiersin.org/articles/10.3389/fnbeh.2025.1516127/full.

Rajeev Gupta, Suhani Gupta, Ronak Parikh, Divya Gupta, Amir Javaheri, and Jairaj Singh Shaktawat. Personalized Artificial General Intelligence (AGI) via Neuroscience-Inspired Continuous Learning Systems, April 2025. URL http://arxiv.org/abs/2504.20109. arXiv:2504.20109 [cs].

Stevan Harnad. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3): 335–346, June 1990. ISSN 01672789. doi: 10.1016/0167-2789(90)90087-6. URL https://linkinghub.elsevier.com/retrieve/pii/0167278990900876.

Borjan Milinkovic and Jaan Aru. On biological and artificial consciousness: A case for biological computationalism. *Neuroscience & Biobehavioral Reviews*, 181:106524, February 2026. ISSN 01497634. doi: 10.1016/j.neubiorev.2025.106524. URL https://linkinghub.elsevier.com/retrieve/pii/S0149763425005251.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision, February 2021. URL http://arxiv.org/abs/2103.00020. arXiv:2103.00020 [cs].

Marcus E. Raichle and Debra A. Gusnard. Appraising the brain's energy budget. *Proceedings of the National Academy of Sciences*, 99(16):10237–10239, August 2002. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.172399499. URL https://pnas.org/doi/full/10.1073/pnas.172399499.

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, January 2016. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature16961. URL https://www.nature.com/articles/nature16961.

Peter Sterling. Allostasis: A model of predictive regulation. *Physiology & Behavior*, 106(1): 5–15, April 2012. ISSN 00319384. doi: 10.1016/j.physbeh.2011.06.004. URL `https://linkinghub.elsevier.com/retrieve/pii/S0031938411003076`.

Leonardo Villalobos-Arias, Grant Forbes, Jianxun Wang, David L Roberts, and Arnav Jhala. Minding motivation: The effect of intrinsic motivation on agent behaviors, 2025. URL `https://arxiv.org/abs/2507.19725`.

Yuqi Xue and Jian Huang. ReGate: Enabling Power Gating in Neural Processing Units. In *Proceedings of the 58th IEEE/ACM International Symposium on Microarchitecture*, pp. 1160–1177, October 2025. doi: 10.1145/3725843.3756038. URL `http://arxiv.org/abs/2508.02536`. arXiv:2508.02536 [cs].