

# Acceleration into Chaos?

## A Systemic Critique of Sector-Neutral AI Acceleration and the $X^\infty$ -Model as an Ethical-Mathematical Counterproposal

The Auctor

`x_to_the_power_of_infinity@protonmail.com`

X: [@tothepowerofinf](#)

GitHub: [Xtothepowerofinfinity](#)

April 22, 2025

### Abstract

This paper critiques the assumptions of the “AI Acceleration: A Solution to AI Risk Policy” paper, which argues that sector-wide acceleration of technological progress is suitable for risk mitigation. We demonstrate that this assumption, particularly in the context of Artificial Intelligence (AI), is dangerously oversimplified due to recursive self-optimization, feedback loops, and chaotic transitions. Drawing on mathematical models, systems theory, and ethical principles embedded in the  $X^\infty$ -Model (an accountability-based governance model with cap logic, feedback obligation, and auditable delegation; project status and preliminary information at: [https://github.com/Xtothepowerofinfinity/Philosophie\\_der\\_Verantwortung](https://github.com/Xtothepowerofinfinity/Philosophie_der_Verantwortung)), we advocate for a feedback-based, responsibility-oriented approach to AI development. Sector-wide acceleration without specific AI risk modeling demonstrably increases the risk of a “wild singularity” rather than reducing it. Available under CC BY-SA 4.0.

## 1 Introduction

The paper “AI Acceleration: A Solution to AI Risk Policy” ([Trammell and Aschenbrenner, 2024](#)) posits that uniform, sector-wide acceleration of technological progress—particularly under regulatory oversight—could mitigate existential risks. The authors assume that shortening risky transition phases (“phase compression”) reduces cumulative risks.

This work counters with a systemic, mathematical, and ethical critique. Especially in the AI domain, acceleration strategies do not act linearly but are self-reinforcing, undermining the paper’s core assumptions. The  $X^\infty$ -Model (in preparation) provides a theoretical and ethically grounded counterproposal.

## 2 Mathematical Critique: The Omission of $R_{\text{AI}}(t)$

The discussed paper views total risk as an additive function:

$$R_{\text{total}}(t) = R_{\text{sectors} \neq \text{AI}}(t) + R_{\text{AI}}(t). \quad (1)$$

The dynamics of  $R_{\text{AI}}(t)$  are assumed to be either constant or proportional to the general development rate  $S_i$ .

This simplification misrepresents the actual risk structure in AI:

$$R_{\text{AI}}(t) = h(S_{\text{AI}}(t), C(t)), \quad (2)$$

where  $S_{\text{AI}}(t)$  denotes the development state and  $C(t)$  the control capacities. With exponential growth of  $S_{\text{AI}}(t)$  and linear or delayed growth of  $C(t)$ ,  $R_{\text{AI}}(t)$  inevitably dominates total risk:

$$R_{\text{AI}}(t) \sim e^{\lambda t}, \quad \lambda > 0. \quad (3)$$

This behavior is identified as a central risk potential of AGI in [Bostrom \(2014\)](#); [Yudkowsky \(2008\)](#); [Russell \(2019\)](#); [Tegmark \(2017\)](#) and is explicitly addressed by the  $X^\infty$ -Model.

## 3 Self-Reinforcement and Wild-Type Singularity

The specific danger of recursive self-optimization is described by a nonlinear feedback differential equation:

$$\frac{dS_{\text{AI}}}{dt} = k \cdot S_{\text{AI}} + \alpha S_{\text{AI}}^2. \quad (4)$$

The solution exhibits blow-up behavior:

$$S_{\text{AI}}(t) \sim \frac{1}{t_0 - t}, \quad (5)$$

indicating that  $S_{\text{AI}}(t)$  diverges to infinity as  $t$  approaches the critical time  $t_0$ .

This behavior not only mathematically describes a singularity but also reflects the real danger that, without sufficient control capacity, AI development dynamics escalate uncontrollably. The term “wild singularity” denotes an unchecked, self-reinforcing development that is neither ethically legitimized nor systemically feedback-controlled.

The  $X^\infty$ -Model thus demands as an ethical and mathematical minimum condition:

$$\frac{dC(t)}{dt} \geq \frac{dS_{\text{AI}}(t)}{dt}, \quad (6)$$

to ensure stability and controllability.

## 4 Systemic Feedback as a Counterproposal

The idea of mitigating risks through feedback is central to the  $X^\infty$ -Model. Control is organized not through centralized planning but through decentralized, auditable feedback mechanisms that document responsibility based on cap history and impact.

The core of the model lies in the conviction that ethical legitimacy must be structurally, not declaratively, grounded: Those who bear responsibility must do so transparently, traceably, and with feedback. Power is not an end in itself but derives from borne accountability.

Accountability replaces power. Cap legitimizes impact, not status.

Thus, the model is not merely a theory but an architecture for operationalizable ethics.

## 5 Ethical Foundation of the $X^\infty$ -Model

The ethical foundation of the  $X^\infty$ -Model radically differs from utilitarian or pragmatic risk governance approaches. While many existing models, including Dewey’s differential technological development, implicitly or explicitly aim to maximize overall utility (or minimize aggregate risks), the  $X^\infty$ -Model is based on an absolutist ethic of protecting the most vulnerable.

This ethic is not the result of a trade-off but the constitutive principle of legitimacy itself. This means:

*No technological development is legitimate if it endangers the most vulnerable—regardless of the potential overall benefit.*

The model explicitly positions itself against utilitarian evaluation logics such as:

$$V_{\text{life}}(t) = f(W(t)), \tag{7}$$

where  $V_{\text{life}}(t)$  denotes the value of a life and  $W(t)$  its resource base. This equation, typical of many economic risk assessment models, is rejected by the  $X^\infty$ -Model.

Instead, it holds:

Impact legitimizes. Status does not. Protection is a duty.

Accountability becomes the sole permissible basis for influence. The cap system quantifies historically borne responsibility as a measure of permissible resource use.

## 6 The Cap System as a Governance Instrument

The cap system of the  $X^\infty$ -Model operationalizes responsibility. It does not describe power limits but the systemically legitimized capacity to bear responsibility. Cap arises solely from proven, borne responsibility—it is neither inheritable, purchasable, nor delegable without feedback.

The system’s core equation is:

$$\text{Cap}_{\text{total}} = \text{Cap}_{\text{Solo}} + \text{Cap}_{\text{Team}} \leq \text{Cap}_{\text{Potential}}. \quad (8)$$

Delegation occurs only if both sender and receiver possess sufficient cap. Vertical delegation is permissible only if:

$$\text{Cap}_{\text{Sender}} \geq \text{Cap}_{\text{Threshold}} \quad \text{and} \quad \text{Cap}_{\text{Receiver}} \geq \text{Cap}_{\text{Delegation-Min}}. \quad (9)$$

A violation of this rule triggers a systemic blockade. The system does not impose punishment but recognizes ineffectiveness as an automatic consequence of feedback refusal.

## 7 Feedback as Systemic Ethics

Feedback replaces traditional morality in the  $X^\infty$ -Model. Not intention, but impact determines legitimacy. Misconduct is not punished but redirected and blocked.

*Not morality, but impact. Not self, but service. Not chaos, but structure.*

Errors are permitted as long as they are fed back. Avoiding feedback, however, is systemic betrayal of the model.

## 8 Comparison with Alternative Governance Approaches

### 8.1 Differential Technological Development (Dewey, 2015) and Its Limitations in the Context of Exponential AI Dynamics

Dewey (Dewey, 2015) formulates a pragmatic approach to risk mitigation in technological development paths with the concept of *differential technological development*. The core idea is to prioritize technologies with a positive safety profile (e.g., AI alignment methods, robust control systems) while deliberately delaying high-risk development lines (e.g., autonomous, uncontrolled AGI). This principle assumes that the pace of technological progress is a controllable parameter and that safety can be achieved through prioritization within existing funding and decision-making systems.

Mathematically, this approach can be formulated as a prioritization problem:

$$\min R(t) = \sum_i w_i \cdot R_i(t), \quad (10)$$

where  $R_i(t)$  represents the risk exposure of individual technologies  $S_i$  at time  $t$ , and  $w_i$  their weighted prioritization. Risk mitigation is thus conceptualized as the aggregate effect of optimized resource allocation.

Although this approach undoubtedly represents an ethical advancement over blind acceleration strategies (Trammell and Aschenbrenner, 2024), it remains within a linear, static view of

technological development. Dewey does not account for the exponential self-reinforcement of recursive AI optimization nor the inherent feedback effects that fundamentally shape AI system dynamics.

## 8.2 The $X^\infty$ -Model: Feedback as Systemic Duty, Ethics as Structure

The  $X^\infty$ -Model presented here addresses this conceptual weakness. It views technological development not as static sequencing but as a dynamic, self-reinforcing system whose stability can only be ensured through *structured feedback obligation*. The central difference from Dewey lies not merely in the *how* of prioritization but in the *whether* of legitimacy itself: No development path is legitimate within the  $X^\infty$  framework unless it is feedback-controlled, auditable, and responsibility-driven.

The model’s mathematical intuition rests on explicit modeling of self-reinforcement:

$$\frac{dS_{\text{AI}}}{dt} = k \cdot S_{\text{AI}} + \alpha S_{\text{AI}}^2, \quad (11)$$

where the blow-up behavior ( $S_{\text{AI}}(t) \sim \frac{1}{t_0-t}$ ) is not merely a theoretical artifact but an empirically plausible dynamic of recursive AI self-improvement. The  $X^\infty$ -Model defines as a necessary condition for risk limitation that control mechanisms  $C(t)$  must keep pace with the growth rate of  $S_{\text{AI}}(t)$ :

$$\frac{dC(t)}{dt} \geq \frac{dS_{\text{AI}}(t)}{dt}. \quad (12)$$

This condition is not optional but systemically and ethically mandatory. Its non-fulfillment constitutes an ethical violation by definition.

The  $X^\infty$ -Model further integrates cyclical misincentives (cobweb theorem, [Ezekiel, 1938](#)) and dynamic interdependencies across sector-wide couplings:

$$R_{\text{total}}(t) = R_{\text{sectors} \neq \text{AI}}(t) + R_{\text{AI}}(t), \quad (13)$$

where  $R_{\text{AI}}(t)$  can dominate once control mechanisms lag behind.

## 8.3 Ethical Singularity: The Absolute Priority of Protecting the Most Vulnerable

A crucial difference between Dewey and  $X^\infty$  lies in their ethical stance: While Dewey’s approach remains utilitarian, focusing on global harm minimization, the  $X^\infty$ -Model rejects any form of utilitarian value weighting. Protecting the most vulnerable is not a means to an end but the ethical endpoint of the system structure itself. There is no justification for acceleration without feedback. There is no argument for prioritization without accountability. The *inevitability* of the  $X^\infty$ -Model is not dogmatic—it is mathematically, system-theoretically, and ethically necessary.

The model is thus not one alternative among many but a necessary foundation for any serious AI risk governance that aspires to more than hope: It structures duty. It operationalizes ethics.

And it immunizes against the systemic excuse that “things couldn’t be done differently at the time.”

## 9 Conclusion

The premises of the “AI Acceleration” paper (Trammell and Aschenbrenner, 2024) do not lead to risk reduction but exacerbate dynamic destabilization. The  $X^\infty$ -Model offers an ethically grounded and mathematically consistent counterproposal that enables stabilization through feedback and responsibility logic. The full publication of the  $X^\infty$ -Model will follow.

## References

- Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.
- Daniel Dewey. Differential Technological Development. Open Philanthropy Project, 2015. <https://www.openphilanthropy.org/research/differential-technology-development/>.
- Mordecai Ezekiel. The Cobweb Theorem. *Quarterly Journal of Economics*, 52(2):255–280, 1938.
- Stuart Russell. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking Press, 2019.
- Michael J. Sandel. *The Case Against Perfection: Ethics in the Age of Genetic Engineering*. Harvard University Press, 2007.
- Max Tegmark. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Alfred A. Knopf, 2017.
- Philip Trammell and Leopold Aschenbrenner. AI Acceleration: A Solution to AI Risk Policy. Global Priorities Institute, GPI Working Paper No. 13-2024. <https://globalprioritiesinstitute.org/wp-content/uploads/AI-Acceleration-A-Solution-to-AI-Risk-Policy.pdf>.
- Eliezer Yudkowsky. Artificial Intelligence as a Positive and Negative Factor in Global Risk. In Nick Bostrom and Milan M. Ćirković (Eds.), *Global Catastrophic Risks*, pp. 308–345. Oxford University Press, 2008.