Meltdown

"In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it."
– **Herbert A. Simon**

By Michael Zargham and Ilan Ben-Meir

In the 20th century, humanity demonstrated that nuclear fission – "splitting the atom" – results in the release of tremendous amounts of energy. In short order, we also discovered two primary use-cases for such energy: It can be *harnessed* to power a reactor, or *unleashed* with the destructive force of a bomb.

At a high level, the difference between these two applications is startlingly straightforward. Nuclear fission sets off a chain reaction; a reactor *regulates* this reaction, while a bomb initiates a "runaway" reaction that produces an explosion.

The development of transformer architectures similarly "split the atom" for natural language processing (and ultimately for content production across modalities – text, speech, image, and video). The increasing availability of "generative" AI has collapsed the attention cost of content creation, setting off a cascade of production that increasingly threatens to spiral out of control.

*The chain reaction has begun. The question is whether or not it can be **regulated**.*

With a prompt, a few clicks, or a simple API call, we can now summon entire documents, artworks, codebases, or business plans out of the void. These outputs are subsequently integrated into the stockpiles of data upon which models are trained, compounding their influence on the broader ecosystem of information production. The bottleneck is no longer generation, it is *judgment*.

Judgment is rate-limiting. In organizations, oversight is a function of human processes (like quality assurance, compliance review, and release control) that still operate at human speeds. Such processes rely on laws and precedents, social norms, tacit knowledge, and informed consent – all of which are brittle when overrun by scale. What were once well-tuned workflows are currently collapsing under the weight of endless drafts, hallucinated references, and polished-but-substantively-incorrect artifacts.

The information environment is going supercritical, and the institutions that have long structured our society are melting down.

In a nuclear reactor, a "meltdown" occurs when the systems that regulate the core reaction fail. This core reaction is remarkably similar to the process that sets off a nuclear bomb – but in a bomb, it is allowed to continue unchecked until it blows its source apart ("disassembly"), while in a reactor, it is stabilized through layers of containment, feedback loops, control rods, trained operators, and protocols that govern how the system *reacts* to each change. Without these regulatory systems, fission's generative potential becomes a destructive force: bombs explode, reactors melt down.

The same principle holds for natural, artificial, and even institutional intelligence.

Insofar as the institutions of our society are melting down in response to the emergence of generative technologies, it is not because we have yet to figure out how to get these technologies to work properly – it is because they work *too well*. The problem is not that the generative power that they unleash is insufficient for our ends, but that its reach exceeds our grasp.

Generative AI produces information so *efficiently* that it has overwhelmed humanity's ability to manage that production *effectively*; consequently, we are experiencing information overload at a societal scale. Left unchecked, generative AI outputs cascade into a distributed denial-of-service (DDOS) attack, overwhelming our institutions with unmanageable (and thus destabilizing) quantities of information. As the disparity between this superabundance of information and the amount of attention available to process it widens, the quality of decision-making degrades.

If generative AI technologies represent the "warheads" or "reactor cores" that set off artificial intelligence's "core reaction" – the release of informational energy – then making these technologies "more powerful" is *the exact opposite* of what is needed to stabilize the systems that they are already threatening to overload. Instead, we need the equivalent of "control rods" – mechanisms for constraint and feedback – as well as failsafes and safe operating procedures. Without such infrastructure, generative AI can only release informational energy in uncontrolled bursts; with it, this same potency can be harnessed to power the pursuit of a particular purpose. At this point, we do not need systems that are *better at producing information* (which is *a means to some end*, not *an end in itself*), but rather systems that can *interpret, coordinate, verify, and validate such production in service of specific aims.*

What we need now is not better generative AI – it's *regulative* AI: artificial intelligence systems designed to monitor, validate, align, or constrain other systems – especially generative ones – in order to enforce rules, constraints, or norms through feedback and oversight.

Regulative AI is not *opposed to* generative AI; instead, it *makes generative systems viable* through the intelligent application of steering, brakes, filters, and correctives that keep them from self-destructing. As generative AI floods the zone with its outputs, regulative AI provides the necessary *structure* for channeling those outputs toward purpose.

Regulation, in this sense, *is* intelligence: not a secondary feature or bureaucratic afterthought, but what allows any complex system – whether a reactor, a brain, a business, or a society – to remain coherent under stress. It is therefore no accident that the history of artificial intelligence begins with, and primarily consists of, regulative rather than generative technologies. The earliest forms of artificial intelligence were not *creative*, but *corrective*; the purpose of these **regulators** (e.g., water-clocks, governors, thermostats, and servomechanisms) was not to simulate brilliance, but to enable engineered systems to **self-stabilize**.

This sort of regulative AI has long enabled systems to be effective. In water-clocks, a system of weights and levers regulated the rate at which water flowed through the mechanism, to ensure that time's passage is being marked at a consistent pace; in steam engines, centrifugal governors throttled the steam valve in order to maintain a steady speed; in industrial automation, analog feedback loops used the difference between measured and desired electrical signals to correct for process drift and error in real-time; in digital control systems, PID controllers combined signals from discrete sensors to keep temperatures, voltages, and trajectories within specified ranges. These systems were all *intelligent*: They sensed, applied logic, and acted (within a narrow but vital scope) to keep some process from drifting out of its range of safe and effective states. What made such systems impressive was not their creative capacity, but their *reliability* – even when faced with noise, delay, and uncertainty.

This pursuit of reliability under changing conditions was the foundation of **cybernetics**, the mid-20th-century science of "control and communication in the animal and the machine." It was here, and not in Large Language Models, that the intellectual project of "Artificial Intelligence" first took shape,[1] and it is this lineage that we must recover if we want to keep today's generative technologies from spiraling out of our control. In its focus on **generation**, modern machine learning has often ignored the necessity of **regulation**. As generative technologies crank out "content" at superhuman speeds, however, cybernetic thinking about feedback, correction, constraint, and coordination takes on a renewed relevance: *It's what keeps systems from melting down.*

The assertion that viability requires regulation is rooted in *biology*, not *ideology*. Biological systems are not considered intelligent because they can *generate* complexity, but because they remain viable in the face of complexity. Survival, as a process, involves *satisfying constraints*, not *maximizing output*; regulation is how biological systems remain within their sustainable operating ranges.

Homeostasis, for example, is the body's thermostat.

The immune system – a distributed, adaptive network capable of detecting threats, remembering past encounters, and coordinating a response – offers a more complex illustration. Its "intelligence" lies in its ability to discriminate: it must distinguish self from non-self, healthy from pathological, threat from noise. If the immune system overreacts, it causes autoimmune disorders; if it underreacts, infection wins. The health of an immune system depends on how well it is able to manage this dynamic balancing, rather than the brute force of the response that it mounts. Similarly, healthy ecosystems regulate populations, flows of energy, and cycles of matter through layered feedback mechanisms such as predator-prey dynamics, nutrient recycling, and ecological niche formation. When these feedback mechanisms are overrun (e.g., by the decline of keystone species), the ecosystem *collapses*.

When regulation fails in biological systems, the results are both destructive and instructive. Cancer is unchecked cellular generation – a destructive runaway process that begins with a single cell losing its ability to respond to internal and external regulatory signals. Ignoring its prescribed life cycle, this cell proliferates without regard for the body's needs, consuming critical resources and overwhelming vital systems until the host organism can no longer function. Invasive species represent a failure of regulation on the *ecological* scale, while pandemics represent simultaneous breakdowns in biological and normative regulation at the *civilizational*

---

[1] Amber Case. (2023). "Inside the Very Human Origin of the Term 'Artificial Intelligence' – And Its Seven Decade Boom/Bust Cycle," *BlockScience* (Medium). Available at: https://medium.com/block-science/inside-the-very-human-origin-of-the-term-artificial-intelligence-and-its-seven-decade-c36e0326245e.

scale; an epidemic becomes a pandemic when it overwhelms biological, medical, and social responses, and self-replicating pathogens are able to reproduce and spread faster than they can be contained and treated.

All of the above are examples of uncontrolled growth – yet each represents a *failure of regulation*, rather than a *triumph of generativity*. Without regulation, generative processes can only end in *excess* or *exhaustion*. Unchecked generativity destroys systems under the weight of *unsupportable abundance*.

Regulation, in other words, is the *price of survival*. This is why human societies are structured around **protocols and institutions** – such as markets, states, firms, and legal systems – that function to coordinate our actions at scale.[2] While these institutions are not artificial in the narrow computational sense of the term, they arise from an interplay between engineering and social evolution. They are *evolved and maintained systems of regulation* – **viable systems**, as defined by cyberneticist Stafford Beer, capable of maintaining identity and purpose in changing environments through feedback, adaptation, and distributed control.

Viable systems are effective intelligences distributed via roles, protocols, policies, and practices. Markets regulate supply and demand through pricing signals – feedback mechanisms that balance distributed generativity. Firms establish policies, escalation paths, and release processes that allow hundreds (or even thousands) of individuals to coordinate actions safely. Legal systems codify social norms as binding constraints, mediating conflicts and preventing the replication of exploitative and predatory behaviors. Scientific institutions provide epistemic regulation (in the form of peer review, reproducibility standards, and editorial control) in order to stabilize knowledge production without losing the capacity to revise established understandings in light of new theories and evidence. Last but not least, peer-production communities generate and maintain open-source software and knowledge repositories, leveraging an assemblage of social and technical protocols to collectively regulate distributed production into useful, stable, and trustworthy outputs.

The above examples are all **artificial organizational intelligences** that must manage tensions between openness and control, creativity and coherence in order to realize their animating purpose of regulating human production, feedback, and decision-making at a super-individual scale.[3] Generative AI can *integrate with* and *complement* these systems, but it cannot *replace* them – and our existing institutions are already straining under the weight of generative technologies. Moderators can't review every AI-generated post or comment online, reviewers

---

[2] Michael Zargham & Ilan Ben-Meir. (2025). "Protocols and Institutions." In *Web3 Blockchain Economic Theory*. Eds. M. Swan, S. Takagi & F. Witte. (Forthcoming, 2025-2026e). London: World Scientific. [2nd Edition of *Blockchain Economics: Implications of Distributed Ledgers*. (2019). Eds. M. Swan, J. Potts, S. Takagi, F. Witte, & P. Tasca. London: World Scientific.]. Preprint available via Zenodo at: https://doi.org/10.5281/zenodo.15122312

[3] Ellie Rennie, Kelsie Nabben, Michael Zargham, Jason Potts, Brooke Ann Coco, Luke Miller, and Matthew Green (2025). "Building the Loop: The Role of Ethnography in Artificial Organisational Intelligence." *2025 Ethnographic Praxis in Industry Conference (EPIC) Proceedings* (forthcoming). Preprint available via SSRN at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5516298

can't check every AI-summarized paper or pull request, and policy teams can't keep up with AI-generated documents that might contradict company policies, violate contracts, or even break laws. We're witnessing a swamping of regulative capacity – not because our institutions have lost the ability to perform their vital societal functions, but because we're asking them to process information with greater volume, velocity, and variety than ever before.

The failure of generative AI is *not* a failure to generate – it's a failure to *validate*. GenAI pilots don't fail because the model is unable to produce outputs; they fail because *the accountable parties **should not trust** those outputs*. Compliance can't approve them, legal can't sign off, and domain experts don't have time to check every sentence for drift, every table for hallucination, and every claim for policy violation. The systems in question *work as designed*, but cannot recognize or correct for misalignment between intent and outcome.

This isn't a creative problem – it's a regulatory one. In the language of engineering, it is about *verification* and *validation*. **Verification** asks: *Did we build the thing right?* It is the domain of specifications, checklists, and formal rules; the kinds of things that machines *can* be trained to do. **Validation**, on the other hand, asks: *Did we build the right thing?* Its domain is that of judgement, relevance, and fitness to purpose; the kinds of things that *require* human discretion and context.

The real bottleneck in deploying AI is not generation, or even verification. It is *validation at scale*: the human act of deciding whether output that has been generated by AI is *actually good enough* for real-world use. Such judgement, which is inherently contextual and often tacit, *cannot* be automated.

Absent adequate regulative mechanisms, the *fluency* of generative systems can easily mask a lack of *fidelity*. Such systems are capable of producing a tremendous volume of outputs that *seem plausibly correct*, but subtly violate some policy, drift from the terms of a contract, overlook a critical exception, or ignore hard-learned lessons. This behavior is characteristic of unregulated **feedforward** systems – they push forward blindly, without calibration on the basis of downstream consequences.

Even with careful planning, therefore, feedforward control always eventually diverges from its objectives. In order to reliably achieve outcomes in complex environments, feedback is required.

Feedback systems can detect when outputs contradict stated goals, violate explicit constraints, or drift from precedent – and can respond with corrective actions. We need systems that can **check, revise, and reject** the outputs that AI generates. *Such systems must also involve humans with the requisite judgement to define what qualifies as "good enough."* This is the domain of regulative AI: signal processing and control systems engineering applied to natural language, in order to regulate the explosive potential of generative AI.

If generative AI is the mechanism through which AI systems "split the information atom," then regulative AI is everything that differentiates a "reactor core" from a "warhead" – the layered systems and processes that *stabilize* the release of informational energy from such fission, and *harness* that energy for safe, purposeful, and productive ends.

Both warheads and reactor cores, it is worth noting, are **engineered systems** built around fission's natural force. The same is true of both generative and regulative AI. An engineered system, by definition, is "a system designed or adapted to interact with an anticipated operational environment to achieve one or more intended purposes while complying with applicable constraints."[4] Crucially, engineered systems *redistribute harms and benefits* – creating leverage (some purposes are enabled and some constraints are enforced), while also demanding trade-offs (certain purposes are enabled *because* certain constraints are enforced). In order to fulfill one set of goals, an engineered system may impinge upon others. There is no truly neutral infrastructure.

Regulative AI does not *resolve* this tension; it *acknowledges* and *accounts for* it.

Every intelligent system is both *influenced by* and *exerts influence on* its **environment**: shaping behaviors, enforcing constraints, and negotiating relationships. The design of AI systems – their models, interfaces, policies, and defaults – are already shifting our relationships with information, and with each other. AI has impacted the ways that we as humans write, code, learn, communicate, and make decisions. These systems are not just *tools* that we wield, but environments that *we* live in and adapt to; we are regulated *by* AI, even as we seek to regulate it.

Regulative AI is about co-regulation amongst diverse intelligences: humans, machines, and beyond.

Rather than imagining regulative AI as some new frontier of algorithmic oversight, we propose returning to the foundations of cybernetics: to signal processing, control theory, and systems engineering – fields developed to **contend with complexity**, not to generate novelty. We must draw insights from control systems in nuclear and chemical engineering, which carefully moderate generative reactions in order to prevent catastrophe; from computer-aided design and structural simulations, which test whether artifacts can hold their shapes under stress before they are ever built; from model-based systems engineering and multidisciplinary design optimization, which coordinate multiple domains and constraints across the lifecycle of complex systems; and from digital twins in aerospace and critical infrastructure, which use real-time data and simulation to monitor, maintain, and adjust deployed systems in the field. These antecedents offer more than apt analogies; they also anchor the concept of regulative AI to working practices in related real-world mission-critical and safety-critical environments. Now we must apply the same level of rigor, foresight, and coordination to the digital infrastructures that are rapidly reconfiguring our world.

---

[4]International Council on Systems Engineering (INCOSE) (n.d.). "Engineered system definition." INCOSE. Available at https://www.incose.org/about-systems-engineering/system-and-se-definitions/engineered-system-definition.

The purpose of Regulative AI is to harness generativity – to make it viable, legible, and accountable as it increasingly integrates into our everyday lives. *Today's landscape is defined by a deficit of judgment, not by inadequate intelligence.* We have built systems that can generate text, code, images, and decisions at previously-unimaginable speeds and scales – but without complementary systems to direct, constrain, validate, and co-regulate these systems, their outputs overwhelm and exhaust us. They consume and fatigue our attention, leaving us lacking both the capacity for judgement and a means of holding its exercise to account.

The future of AI will be defined by **more purposeful coordination, made possible through complementary generative and regulative intelligences.** It will be inaugurated by the engineering of artificial organizational intelligences that both *produce* and *adapt* – systems that don't just scale, but can verifiably satisfy constraints and demonstrably pursue their stated purposes *at scale*. It will be realized by AI systems that are not only powerful, but also trustworthy, testable, corrigible, and tailored to work reliably within operating conditions as messy and open-ended as the life-worlds in which we humans labor and live.

We will not reach this future, however, unless we can avert the present meltdown. This is the work of regulative AI.

****

Glossary

Artificial Intelligence: Engineered systems – typically computational – that process information, learn from data, and adapt their behavior to achieve goals in digital or physical environments. **Cf.** Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Artificial Organizational Intelligence (AOI): "The capacity for organisations to make their knowledge (including routines, practices, rules, boundaries, and tacit know-how) legible and governable. AOI is not about replacing organisations with AI, but about enabling them to talk to themselves and to each other through regulated feedback loops." **Cf.** Ellie Rennie, Kelsie Nabben, Michael Zargham, Jason Potts, Brooke Ann Coco, Luke Miller, and Matthew Green (2025). "Building the Loop: The Role of Ethnography in Artificial Organisational Intelligence." *2025 Ethnographic Praxis in Industry Conference (EPIC) Proceedings* (forthcoming). Preprint available via SSRN at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5516298.

Constraint Satisfaction: A process by which systems select or evaluate behaviors that meet predefined limitations, rules, or environmental conditions. **Cf.** Wikipedia, https://en.wikipedia.org/wiki/Constraint_satisfaction_problem.

Control Engineering: The engineering discipline that applies control theory to design and implement systems – mechanical, electrical, biological, or institutional – that measure outputs and use corrective feedback to ensure desired behavior or stability. **Cf.** Wikipedia, https://en.wikipedia.org/wiki/Control_engineering.

Cybernetics: "The transdisciplinary study of circular causal processes such as feedback and recursion," concerned with general principles across engineering, ecological, economic, biological, cognitive, and social systems. **Cf.** Wikipedia, https://en.wikipedia.org/wiki/Cybernetics

Engineered System: "A system designed or adapted to interact with an anticipated operational environment to achieve one or more intended purposes while complying with applicable constraints." **Cf.** INCOSE (International Council on Systems Engineering), https://www.incose.org/about-systems-engineering/system-and-se-definitions/engineered-system-definition.

Feedback: Processes in which a system's output is routed back as input in order to compare actual performance to goals or desired states, enabling correction, stability, or adaptation. **Cf**. Wikipedia, https://en.wikipedia.org/wiki/Feedback.

Feedforward: Processes in which signals or inputs are transmitted through a system to affect its outputs, with little or no feedback loop modifying the behavior based on the outcome. **Cf.** Wikipedia, https://en.wikipedia.org/wiki/Feedforward_(control).

Generative: Possessing the capacity to enable unanticipated, unfiltered innovation or change – especially from diverse, widely distributed contributors. **Cf.** Zittrain, J. (2008). *The Future of the Internet and How to Stop It*. Yale University Press.

Generative AI: "A subfield of artificial intelligence that uses generative models to produce text, images, videos, or other forms of data. These models learn the underlying patterns and structures of their training data and use them to produce new data based on the input, which often comes in the form of natural language prompts." **Cf.** Wikipedia, https://en.wikipedia.org/wiki/Generative_artificial_intelligence

Infrastructure: "The assets, systems, and networks, whether physical or virtual, whose incapacitation or destruction would have a debilitating effect on security, national economic security, public health or safety, or any combination thereof." **Cf.** Guide to the Systems Engineering Body of Knowledge, https://sebokwiki.org/wiki/Infrastructure_(glossary).

Institution: “An enduring pattern of behavior that remains stable (in the sense of Lyapunov) over time, indicating that the system maintains certain properties despite variations in specific situations and individual actions.” **Cf.** Michael Zargham & Ilan Ben-Meir. (2025). “Protocols and Institutions.” In *Web3 Blockchain Economic Theory*. Eds. M. Swan, S. Takagi & F. Witte. (Forthcoming, 2025-2026e). London: World Scientific. [2nd Edition of *Blockchain Economics: Implications of Distributed Ledgers*. (2019). Eds. M. Swan, J. Potts, S. Takagi, F. Witte, & P. Tasca. London: World Scientific.]. Preprint available via Zenodo at: https://doi.org/10.5281/zenodo.15122312.

Intelligence: The capacity of a system – biological, artificial, or institutional – to process information, adapt behavior, and achieve goals effectively across diverse environments. **Cf.** Legg, S., & Hutter, M. (2007). “A Collection of Definitions of Intelligence.” Frontiers in Artificial Intelligence and Applications, 157, 17–24.

Natural Intelligence: Adaptive behaviors and information-processing capacities that emerge through evolutionary or ecological processes in biological systems. **Cf.** Sterelny, K. (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. Blackwell Publishing.

Natural Language Processing: “Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications.” **Cf.** Liddy, E.D. (2001). “Natural Language Processing.” In *Encyclopedia of Library and Information Science,* 2nd Ed. Marcel Decker, Inc.

Protocol: “A set of humanly constructed rules, norms, or shared strategies that specify a subset of actions available to individuals in a given situation, thereby coordinating individual behavior.” **Cf.** Michael Zargham & Ilan Ben-Meir. (2025). “Protocols and Institutions.” (See **“Institution”** [above] for full citation).

Regulation: The process by which a system maintains certain essential variables within acceptable bounds in the face of disturbances. A regulator must have enough variety (i.e. enough different internal “states” or responses) to counteract external disturbances. If the environment (or disturbances) can push the essential variables into many possible states, the regulator must be able to respond with at least as many distinguishable responses; otherwise some disturbances will “get past” the regulator and upset the system. **Cf.** Ashby, W.R. (1956). *An Introduction to Cybernetics*. Wiley.

Regulative: Possessing the capacity to constrain or shape behavior – whether intentionally or unintentionally – enabling systems to maintain coherence, enforce norms, or limit risks. **Cf.** Lessig, L. (1999). *Code and Other Laws of Cyberspace*. Basic Books.

Regulative AI: Artificial intelligence systems designed to monitor, validate, align, or constrain other systems – especially generative ones – in order to enforce goals, rules, constraints, or norms through feedback and oversight. **Cf.** This essay.

Signal Processing: The engineering discipline concerned with “analyzing, modifying, and synthesizing signals” (such as sound, images, or sensor data) to extract information, reduce noise, and render sense from raw data. **Cf.** Wikipedia, https://en.wikipedia.org/wiki/Signal_processing

Validation: Evaluation of whether a system is fit for purpose in its operational context – whether the outcomes are good enough – recognizing that adequacy cannot be fully reduced to a finite checklist and therefore requires judgment. **Cf.** Guide to the Systems Engineering Body of Knowledge, https://sebokwiki.org/wiki/Validation_(glossary).

Verification: Evaluation against a completely specified set of rules, models, or criteria to confirm a system has been built correctly and conforms to its specification. **Cf.** Guide to the Systems Engineering Body of Knowledge, https://sebokwiki.org/wiki/Verification_(glossary).

Viable System: An organizational system – such as a firm, bureaucracy, or state – that maintains its identity and functionality through self-regulation and adaptive feedback across internal subsystems and external environments. **Cf.** Beer, S. (1979). *The Heart of Enterprise.* Wiley.

Note: The definitions above are either quoted directly from the provided citations, or paraphrased for clarity and consistency with the topic of this post.