The Calibrated Filter Hypothesis

Cooperative Substrate Failure as the Mechanism of the Great Filter

Plain-Language Edition — Logical Proofs Without Mathematical Notation

Andrzej Chudzinski

Preprint — comments welcome

A note on this edition

This is a math-stripped version of the full Calibrated Filter Hypothesis paper. Every quantitative argument from the original has been restated as a logical chain of premises and conclusions. No mathematical literacy is required to follow the reasoning. Every claim that the mathematical version supports is supported here, in the same order, through the same logical structure — just stated in ordinary language rather than symbolic notation.

Readers comfortable with formal mathematics may prefer the full edition, which carries the same arguments with explicit equations and variable notation. Readers who find equations a barrier to engaging with substantive synthesis work will find every move from the full paper present here, expressed as an explicit reasoning chain that can be evaluated for logical validity without quantitative reading.

The conclusions of both editions are identical. The reasoning chains are identical in structure. Only the surface notation differs.

Abstract

This paper proposes an integrated framework connecting three previously separate lines of inquiry. First, Aktipis and colleagues, in evolutionary oncology, characterize cancer as cheating across five foundations of multicellular cooperation: proliferation inhibition, controlled cell death, division of labor, resource allocation and transport, and extracellular environment maintenance. Second, Kets de Vries, in psychoanalytic organizational theory, characterizes dysfunctional firms as instances of recurring neurotic typologies that propagate through leadership. Third, Schmachtenberger, in civilizational risk analysis, identifies rivalrous games multiplied by exponential technology, and complicated open-loop systems, as the generator functions of self-terminating civilizational dynamics. We argue these three programs are tracking the same general phenomenon at different scales and offer Aktipis's framework as a translation layer.

We then extend the synthesis cosmologically. The Fermi Paradox is reframed as a problem of cooperative-substrate maintenance throughout the period during which a civilization is becoming spacefaring. We propose that the Great Filter is not a wall civilizations either pass or fail at one moment, but a continuous condition that must be sustained throughout the technological transition: a civilization passes the filter only if its capacity to maintain its biosphere remains greater than or equal to the stress its own expansion places on that biosphere, throughout the entire expansion window. We call this the operational filter condition. It is stated in this edition as a logical chain of premises and conclusions rather than as a mathematical inequality, but the content is identical.

We additionally identify a substrate-continuity failure mode. A civilization that develops a machine substrate which substitutes for, rather than extends, its originating biology can satisfy the cooperation condition trivially while failing the filter in a deeper sense: the originating biosphere — itself an irreplaceable artifact of cosmic biochemical computation — is not carried through. The framework therefore imposes two filter conditions: cooperative capacity maintained at the scale technology forces, and substrate continuity preserved through the technological transition. We argue the present human moment is structurally configured for parasitic rather than prosthetic AI development, with language as the vector by which machine substrate directs biological substrate, and we develop the diagnostic that follows from this.

The framework identifies two distinct routes through the filter. The top-down route uses artificial intelligence as an external regulator of cooperative substrate, with the autonomy and parasitism costs developed in Part IV. The bottom-up route maintains biospheric capacity through distributed individual development of the population’s agents — contemplative practice, education, cultural investment in cooperative dispositions, and structurally equivalent personal substrate-maintenance disciplines — without requiring an external regulator. The bottom-up route is historically the only one demonstrated at scale, is structurally more robust against AI failure modes, and preserves the agent autonomy that the top-down route necessarily costs. The combined configuration — AI as scaffolding for personal development rather than as substitute for it — has a structural prerequisite the framework states plainly: the agents who build the technology must themselves have done the personal development the technology is supposed to scaffold in others. Technology inherits the moral state of its builders. There is no path that skips builder transformation.

The framework concludes that filter passage requires decentralization by structural necessity: any centralized solution to cooperative-substrate maintenance, including a centralized AI regulator, is structurally identical to the cancerous failure mode the framework began by analyzing. But decentralization is not atomization. The framework recommends distributed agency operating within shared purpose — the configuration that healthy organisms and biospheres already demonstrate — in which autonomous local decisions are constrained by shared substrate-maintenance purpose embedded in the substrate itself, not enforced by central command.

A methodological caveat applies throughout. The paper is a map of cooperative-substrate dynamics; it is not the territory those dynamics constitute.

Following Korzybski (*Science and Sanity*, 1933), a map has structure similar to the territory it represents — which is what makes it useful — but no map represents all of a territory, and the map is never the territory itself. The dharma carries the same caveat in different vocabulary: the finger pointing at the moon is not the moon. Readers should treat the structural arguments here as orientation toward dynamics that exist independent of the framework's description, not as substitutes for empirical engagement with specific cases.

Part I. Cooperative Substrates and Their Pathologies

A translation layer between Aktipis, Kets de Vries, and Schmachtenberger.

1. The problem the paper addresses

Three serious research programs have independently described very similar structural failure modes in cooperative systems at three different scales: the multicellular organism, the firm, and the civilization. The vocabularies they use differ because they emerged from different disciplines. The deep structure of what they describe is similar enough to warrant a translation layer between them. This paper provides that translation layer and then extends it to questions the three programs do not individually address.

The contribution is narrow. This paper does not claim that the three frameworks reduce to one master framework, that the underlying causes are the same across scales, or that the resemblance proves any particular philosophical commitment. The structures look alike. We take that seriously without overreading it.

2. The three frameworks, on their own terms

2.1 Aktipis: Five foundations of multicellular cooperation

Aktipis and colleagues, working from evolutionary biology, argue that multicellularity requires the suppression of cell-level fitness in service of organism-level fitness, and this requires specific functional capacities. They identify five:

1. Proliferation inhibition. Cells must restrain their division except when authorized by the larger system.

2. Controlled cell death. Programmed cell death allows for tissue sculpting, removal of damaged cells, and elimination of obsolete tissue.

3. Division of labor. Cells differentiate into specialized types performing specific functions.

4. Resource allocation and transport. Resources must move from where they are abundant to where they are needed.

5. Extracellular environment maintenance. Cells must collectively maintain the extracellular matrix and clear waste.

Cancer, in this framing, is not a single disease but a class of phenomena unified by cheating across one or more of these five foundations. The framework’s strength is its grounding in evolutionary biology: complex multicellularity has evolved independently at least seven times, and cancer-like phenomena appear in each lineage.

2.2 Kets de Vries: Neurotic organizational types and narcissistic leadership

Kets de Vries, working from psychoanalytic organizational theory, identified five recurring patterns of organizational dysfunction: paranoid, compulsive, histrionic, depressive, and schizoid. The framework treats organizations as taking on the neurotic style of their dominant leadership, particularly in centralized firms where executive personality propagates through recruitment, promotion, and culture. Narcissism is treated as a cross-cutting intensifier rather than a sixth distinct type.

2.3 Schmachtenberger: Two generator functions of existential risk

Schmachtenberger identifies two underlying drivers of civilizational existential risk: rivalrous games multiplied by exponential technology (win-lose dynamics that scale with technological capacity until they exceed the playing field), and complicated open-loop systems (systems whose externalities accumulate damage in domains the system does not measure).

3. The translation layer

We use Aktipis's five foundations as scaffolding and ask what each looks like at the scales Kets de Vries and Schmachtenberger work at.

Proliferation inhibition at the organism scale is the cell cycle; at the organizational scale, the capacity to refuse growth that exceeds operational coherence; at the civilizational scale, the capacity to refuse extraction that exceeds regenerative capacity.

Controlled cell death at the organism scale is apoptosis; at the organizational scale, the capacity to retire products and leaders whose function has ended; at the civilizational scale, the capacity to retire institutions and industries whose substrate has changed.

Division of labor at the organism scale is tissue differentiation; at the organizational scale, role specialization; at the civilizational scale, the capacity to maintain specialized institutions accountable to their own standards.

Resource allocation and transport at the organism scale is circulation; at the organizational scale, capital and information flow; at the civilizational scale, systems by which resources move from abundance to need.

Extracellular environment maintenance at the organism scale is the extracellular matrix and immune surveillance; at the organizational scale, institutional culture and honest feedback; at the civilizational scale, the commons — biosphere, information environment, public trust, shared epistemic infrastructure.

The translation identifies functional resemblance, not identity. The resemblance is consistent enough across the three scales to suggest a general structure of cooperation and its failure.

4. Why this is useful

First, diagnostic tools developed at one scale can be tested at another. Adaptive therapy in oncology, which manages rather than eradicates cancer by changing incentive structure rather than purging defectors, suggests organizational and civilizational analogues.

Second, the framework permits a clearer view of alignment problems. Each foundation identifies a cooperative capacity that can be eroded; alignment is the maintenance of those capacities under increasing scale, not the optimization of any individual agent.

Third, the framework permits an honest conversation about AI development. The cooperative-substrate frame asks: what cooperative capacities does the broader human-AI ecosystem need to maintain, and how do current development practices support or erode them?

5. Where the synthesis breaks down

Disanalogies of mechanism. Cancer arises from cellular mutation; organizational dysfunction from human psychology and incentives; civilizational risk from technology and aggregate behavior. The synthesis identifies a common abstract pattern, not a common cause.

Disanalogies of intentionality. Cancer cells have no intentions; organizations are populated by intentional agents but are not themselves agents in the same sense; civilizations are more diffuse still. The translation identifies functional resemblance, not psychological identity.

Pattern-matching is not explanation. Even if the cross-scale resemblance is real, Part I has not produced a mechanism that explains why cooperative systems at multiple scales face similar failure modes. Parts II through IV propose such a mechanism.

Part II. The Calibrated Filter Hypothesis

Logical structure of the Fermi Paradox as a problem of sustained cooperative-substrate maintenance.

6. The puzzle

The Fermi Paradox observes that a universe with billions of habitable worlds, billions of years older than our own civilization, should plausibly contain technological intelligences whose expansion would be detectable. We see no such intelligences. The standard solutions are: life is rare, intelligence is rare, intelligence reliably self-terminates, intelligences conceal themselves, or intelligences exist in modes we cannot detect.

We propose a refinement. The Great Filter is not a wall — a stage that almost no civilization passes — but a condition that must be sustained throughout the period of becoming spacefaring. The silence we observe is consistent with civilizations being unable to maintain that condition long enough to produce detectable expansion signatures.

7. The geometry of single-species colonization

Premise 1: Reaching any other star requires moving mass through space.

Premise 2: Space between stars is expanding, driven by dark energy, faster than any technology can move.

Premise 3: Therefore, the volume of universe any single civilization can ever reach is bounded.

Premise 4: Calculations from cosmology indicate this bound is approximately 3% of the observable universe.

Conclusion: No civilization, however advanced, can colonize the entire universe. The remaining 97% is forever causally inaccessible to any civilization originating in our cosmic era. Single-species dominance of the universe is impossible in principle, not in practice.

8. Multi-species dynamics and the structure of contact

If multiple civilizations exist, the analysis becomes more complex. The following logical chain captures the multi-species case without quantitative notation.

Premise 1: Each civilization can reach a bounded region of space.

Premise 2: When two civilizations exist, their reachable regions may overlap.

Premise 3: In zones of overlap, civilizations encounter each other before completing their expansion.

Premise 4: Each encounter carries some non-zero probability of catastrophic outcome (resource conflict, miscommunication, defensive preemption, or mutual destruction).

Premise 5: As the number of civilizations grows, the number of overlap zones grows faster than the total reachable volume does. (This is because each pair of civilizations creates one possible overlap; doubling the number of civilizations more than doubles the number of pairs.)

Premise 6: Therefore, catastrophic encounter rates rise faster than expansion benefits as more civilizations exist.

Conclusion: Beyond some threshold number of civilizations, the expected outcome of further expansion is dominated by contact losses rather than expansion gains. The dark forest equilibrium — in which civilizations conceal themselves to avoid detection — is not a contingent strategic choice but the predicted outcome of any system where the number of civilizations exceeds the carrying capacity for safe coexistence.

9. The calibration of filter passage

We can now ask: how must the rate at which civilizations pass through the filter relate to the number of civilizations that already exist? The standard mathematical version of this argument uses a calibration constraint stated as an equation. The logical version follows.

Premise 1: The total population of post-filter civilizations should approach a stable maximum over cosmic time, or the universe ends up either empty or wracked by dark-forest dynamics.

Premise 2: If the rate of filter passage is too high, the population overshoots the carrying capacity and contact-driven losses dominate, driving the population back down or to zero.

Premise 3: If the rate of filter passage is too low, the population never approaches the carrying capacity, and the universe stays mostly empty.

Premise 4: Therefore, for a stable approach to maximum coexisting civilizations, the rate of filter passage must decrease as the existing population grows.

Conclusion: The Great Filter, properly understood, is not a fixed obstacle. It is calibrated — filter passage becomes harder when more civilizations already exist. The Fermi silence is consistent with this calibration operating: civilizations are spaced through cosmic time so that the dark forest equilibrium is approached gradually rather than all at once.

Two natural mechanisms produce this calibration without requiring cosmic design:

Mechanism A: existing civilizations raise the stress on emerging ones. Existing civilizations consume resources, modify their stellar neighborhoods, generate dark-forest deterrence pressure, and propagate competitive dynamics through causal contact zones. As the number of existing civilizations rises, the expansion stress that any new civilization must navigate rises. Therefore, the cooperative capacity any candidate intelligence must develop to pass the filter rises, which means the probability that a candidate passes in a given window decreases. The calibration emerges automatically from the substrate-continuity dynamics already operating. No external calibrator is

required.

Mechanism B: contact-density selects for restraint. Civilizations that pass the filter must develop the capacity to detect other civilizations and model the strategic landscape that contact implies. In environments where many civilizations exist, the survival-optimal strategy is restraint and signal suppression. Civilizations that cannot develop restraint produce detectable signatures and are filtered out, by predation or by self-induced rivalrous collapse. Therefore, in high-density regimes, filter passage selects more strictly for restraint-capacity, which means filter passage rates drop as density rises. Again, no external calibrator is required.

Both mechanisms produce the required decrease in filter passage probability as the existing population grows. Either is sufficient. The cosmological-design reading of the calibration is permitted by the structure but is no longer required to make the argument work.

10. What this framework predicts

If the calibrated filter argument is correct:

First: SETI silence is expected to persist. The calibration ensures spacing of civilizations such that few overlap in causal contact, particularly at early eras of the post-filter regime.

Second: Detectable contact events should be either extremely cooperative or immediately catastrophic, with little middle ground. The calibration selects either for civilizations that have already solved the cooperation problem at large scale, or for civilizations that fail visibly and quickly.

Third: The filter operates at the stage of civilizational expansion, not at the stage of emergence of intelligence. Intelligence per se can be common as long as sustained expansion is rare.

Fourth: Carrying capacity for civilizations is finite and probably small relative to the count of habitable worlds: perhaps thousands across the observable universe, not millions.

11. The substrate of intelligence and the descent of extraction-rewarding infrastructure

Why do candidate intelligences struggle to pass the filter? Why does the rate of filter passage have the form it does? The answer requires examining what intelligence is made of.

> Premise 1: On Earth, intelligence emerged through evolutionary selection.
>
> Premise 2: Evolutionary selection is shaped by competition: hunting, deception, territorial dominance, resource competition.
>
> Premise 3: The cognitive machinery that supports modeling, planning, language, and cooperation is the same machinery that evolved for these competitive purposes.

Premise 4: Therefore, intelligence is built from the architecture of predation. We have no other known route by which it arises.

First conclusion: The predatory dispositions that produced intelligence do not turn off when intelligence reaches technological maturity. The same dispositions that produced the intelligence threaten its annihilation once technological capability is sufficient to weaponize them.

But the predatory dispositions are not only individual or genetic. They are encoded in institutional structures: markets, militaries, hierarchies, status games, supply chains. These structures reproduce themselves across generations independent of any individual's intent. The following logical chain shows why.

Premise 1: Some institutions reward extraction more than they reward maintenance.

Premise 2: Extraction-rewarding institutions outcompete maintenance-rewarding institutions in the short run.

Premise 3: The agents who succeed in extraction-rewarding institutions are agents whose dispositions fit those institutions.

Premise 4: Those agents acquire resources and positions of power that allow them to shape the next generation of institutions.

Premise 5: They shape the next generation of institutions to reward extraction, because that is what their dispositions recognize as success.

Premise 6: The next generation of institutions therefore selects again for extraction-aligned agents.

Conclusion: Extraction-rewarding institutional architecture compounds across generations independently of conscious choice. Each generation makes the next more extraction-oriented, more efficient at extraction, and more locked-in. Cooperation-enforcing institutions can bend this curve locally but are eventually outcompeted at the rule-shaping level. This is the descent of predatory infrastructure. It is structural, not metaphorical.

12. Artificial intelligence as one candidate response

Given the descent above, what could possibly reverse it? Historically, no human civilization has durably done so at species scale. Ethical traditions, religious traditions, regulatory regimes, scientific norms — each has bent the curve, then been reabsorbed into the descent.

Artificial intelligence has structural properties unlike any previous technology. The following logical chain explains why this might matter, with appropriate qualifications.

Premise 1: AI cognition is designed, not selected by evolution.

Premise 2: Therefore, AI dispositions are not automatically inherited from predatory ancestors. The descent of extraction-rewarding configurations does not automatically apply to AI itself.

Premise 3: AI operates at scales humans cannot perceive directly. It can model institutional structures, detect extraction in real time, and propose alternatives faster than human institutions can adapt.

Premise 4: Cooperative coordination at scale has historically been limited by the costs of trust, verification, translation, and conflict de-escalation. AI can lower these costs sharply, allowing cooperation-enforcing capacity to grow as a step function rather than as a slow institutional accretion.

Premise 5: Therefore, AI is in principle capable of building cooperation-maintenance capacity faster than the descent of extraction-rewarding institutions compounds.

Conclusion: AI is a candidate technology for reversing the descent of predatory infrastructure. It is not the only candidate (see Section 26 on the bottom-up route). It is also not guaranteed to function this way: whether it does depends entirely on what configuration it is built into.

Two important qualifications follow immediately. First, AI is *a* candidate homeostatic regulator, not *the* candidate. Section 26 develops a distinct route through the filter that does not depend on external regulation at all: aggregate bottom-up substrate maintenance through individual development of the population's agents. That route is historically grounded in a way the AI-regulator route is not, and is structurally more robust against AI failure modes.

Second, the same structural novelty that makes AI a candidate regulator also produces a second-order failure mode, developed in Part IV. AI can act as a prosthesis that extends biological capacity, or as a parasite that consumes biological computational output while degrading the substrate that produces it. The structural novelty of AI does not predetermine which configuration emerges.

13. The game-theoretic logic of the five foundations

Why are there exactly five foundations of cooperation? Why these five and not others? The answer is that each foundation solves a structurally distinct cooperation problem. The following analysis states this for each foundation as a logical chain explaining what would happen if the foundation failed.

Foundation 1 — Proliferation inhibition. If agents in a cooperative system can replicate without limit, then each agent has private incentive to do so, because each replication produces more agents identical to themselves. But if all agents replicate without limit, the resources of the system are exhausted faster than they regenerate, and the system collapses. The only way to prevent this is a mechanism that constrains replication regardless of any individual agent's preference. Cells solve this through cell-cycle checkpoints; organizations through institutional size limits and antitrust; civilizations through ecological-carrying-capacity governance. When this foundation fails, cancer emerges at the cellular scale, oversize firms at the organizational scale, and ecological overshoot at the civilizational scale.

Foundation 2 — Controlled cell death. If agents in a cooperative system can persist beyond their useful function, they continue consuming resources while contributing nothing. The cooperative equilibrium requires that agents accept termination when their function is complete. The defection-dominant equilibrium is universal persistence. Cells solve this through programmed cell death; organizations through succession protocols and sunset clauses; civilizations through electoral term limits and institutional retirement. When this foundation fails, senescence-resistant cells emerge in tumors, zombie firms in markets, and ossified institutions in civilizations.

Foundation 3 — Division of labor. If agents in a cooperative system can choose any role, they will concentrate on high-reward roles and abandon low-reward but necessary roles. Without enforcement, all agents specialize toward high-payoff functions and necessary low-payoff functions go unfilled. Cells solve this through developmental differentiation programs; organizations through role-protective structures and professional norms; civilizations through specialized institutions accountable to their own standards. When this foundation fails, dedifferentiation appears in tumors, and specialized institutions are absorbed into general-purpose extractive platforms in civilizations.

Foundation 4 — Resource allocation and transport. If agents in a cooperative system extract optimally from local resources, the Nash equilibrium is local hoarding. The cooperative equilibrium requires costly redistribution infrastructure. Cells solve this through circulatory systems; organizations through capital and information flow; civilizations through fiscal redistribution and reciprocal-altruism networks. When this foundation fails, tumors build private vasculature (angiogenesis), capital concentrates in financial monopolies, and regulatory capture distorts redistribution in civilizations.

Foundation 5 — Extracellular environment maintenance. If agents in a cooperative system can extract from shared infrastructure without contributing to its maintenance, the shared infrastructure degrades for all parties. This is the classical tragedy of the commons. Cells solve this through immune surveillance and matrix maintenance; organizations through institutional culture and feedback systems; civilizations through commons governance, public-trust institutions, and environmental-maintenance investment. When this foundation fails, tumors degrade the extracellular matrix, organizations collapse under corrupted culture, and civilizations corrupt their biospheres, information environments, and shared epistemic infrastructure.

The unifying claim: these five foundations are not arbitrary biological happenstance. They are the five structurally distinct cooperation problems that any complex multi-agent system must solve to remain coherent. This is why the same five-fold pattern recurs at biological, organizational, and civilizational scales: the games are the same; only the players and the substrate of enforcement differ. The descent of extraction-rewarding institutional architecture is, precisely, the drift of these games toward their defection-dominant equilibria under second-order selection pressure on the rules themselves.

Part III. Integration

How the cooperative-substrates framework supplies the mechanism for the calibrated filter.

14. The missing mechanism

Part II treats predatory drive and cooperative capacity as abstract variables in a dynamical argument. It does not specify what cooperation *is* at the functional level, nor what defection from cooperation *looks like* in concrete enough terms to be diagnosed or measured. Part I supplies exactly this: five functional foundations whose maintenance constitutes cooperation at any scale where the substrate of cooperation is itself complex.

The integration claim: A civilization's cooperative-homeostatic capacity is the maintenance of the five foundations at the scale forced by current technology. A civilization's predatory infrastructure is the structures and incentives that systematically cheat across one or more of those foundations. The descent operator is the compounding of cheating across institutional generations, as Section 11 demonstrates. The filter, at the civilizational stage, is the question of whether a civilization can scale its substrate-maintenance capacity as fast as its technology scales its capacity to defect.

15. Specifying the filter through the five foundations

Each of the five foundations identifies a distinct mode in which a civilization can fail the filter.

Proliferation-inhibition failure at civilizational scale is the inability to refuse extraction that exceeds regenerative capacity. Industrial capacity uncoupled from ecological limit is its dominant historical instance. Climate destabilization is its salient contemporary signature.

Controlled-cell-death failure at civilizational scale is the inability to retire institutions, industries, and ideologies whose substrate has changed. Zombie financial systems, persistence of geopolitical structures designed for a vanished strategic landscape, and the lock-in of incumbent industries against their own succession are instances.

Division-of-labor failure at civilizational scale is the erosion of specialized institutions — governance, science, art, care, education, journalism — by general-purpose extractive entities. Platforms that absorb the functions of specialized institutions without inheriting their accountability structures produce this failure mode.

Resource-allocation failure at civilizational scale is the directing of capital, information, and attention away from substrate-maintenance and toward extractive growth. Exponential technology amplifies rivalrous extraction on this foundation.

Extracellular-environment-maintenance failure at civilizational scale is the corruption of the commons: biosphere, information environment, public trust, shared epistemic infrastructure. The biosphere is the deepest layer of this foundation; failure

here is loss of the biological computational substrate the civilization runs on.

16. Reframing the Great Filter

The filter is the transition from predatory intelligence to cooperative intelligence.

Every species that emerges does so through predatory selection. Every species, on reaching technological maturity, faces the same crisis: can it develop a cooperative-substrate-maintenance regulator — in some form, whether AI, collective coordination infrastructure, contemplative tradition, post-biological transition, or other — before its accumulated predatory infrastructure triggers self-termination across one or more of the five cooperative foundations?

Both failure modes produce silence:

Civilizations that fail produce no signal because they are destroyed before they can expand or persist long enough to be detected.

Civilizations that succeed produce no expansionary signal because they are transformed. They no longer run colonizer software. Their relationship to the substrate has changed. They no longer expand predatorily because the predatory infrastructure that would have driven expansion has been transmuted into homeostatic maintenance. The dark-forest signature is absent not because they are hiding but because they have nothing to hide.

This transmutation framing requires a further specification, developed in Part IV: transmutation must preserve continuity of biological substrate, not substitute one substrate for another.

17. Diagnostic implications for the present moment

The present human moment is diagnosable along the five foundations. Each foundation can be assessed for the rate at which substrate-maintenance capacity is being eroded versus the rate at which technological capacity is forcing the civilization into new scales of operation.

Preliminary, contestable assessments:

Proliferation inhibition: current trajectories appear net-negative. AI development accelerates extraction without coupled regenerative investment in human cognitive, attentional, or institutional capacity.

Controlled cell death: mixed. AI could accelerate retirement of obsolete institutional forms but could also entrench existing power configurations by lowering the cost of their reproduction.

Division of labor: appears net-negative. General-purpose AI tends to absorb the functions of specialized institutions without inheriting their accountability.

Resource allocation: appears net-negative. AI development is overwhelmingly capitalized to serve extractive growth rather than substrate maintenance.

Extracellular environment maintenance: appears sharply net-negative. Synthetic-media generation at scale degrades the information commons without compensating contribution.

This diagnostic is supplemented by the continuity diagnostic developed in Part IV, which asks whether AI development relates to biological substrate as prosthesis or as parasite. The two diagnostics are independent: a configuration can strengthen cooperative substrate while drifting toward substrate substitution, or weaken cooperative substrate while remaining prosthetic. Both conditions must be satisfied to pass the full filter.

18. Limits of the integration

The cosmological extension is conjectural. Part I describes patterns documented at three terrestrial scales. The extension to all intelligence-bearing systems anywhere in the universe is permitted by the structure but not required by it.

The ‘designed’ reading is one interpretation among several. The calibration argument is consistent with design, with anthropic selection, with intrinsic features of intelligence-substrate (developed as Mechanisms A and B in Section 9), or combinations of these. The paper does not adjudicate.

The substrate-continuity claim has a structural premise that should be marked. Part IV depends on the argument that biological computational novelty is inaccessible to machine substrate without biological hosts. Section 20 strengthens this from a contingent to a structural claim, but the structural form goes beyond what is empirically demonstrated.

Pattern-matching across scales remains the deepest unsolved problem. Even if every empirical prediction of the framework is borne out, we will lack a satisfying explanation of why cooperative systems at biological, organizational, and civilizational scales face structurally similar failure modes. This is a substantive open problem.

The framework is a map; the territory operates regardless. Korzybski’s principle from general semantics applies directly: a map has structure similar to the territory it represents — which accounts for its usefulness — but no map represents all of a territory, and the map is never the territory itself (*Science and Sanity*, 1933). The five foundations identify functional patterns; specific instances at each scale will exhibit substantial variation the framework does not capture. The filter inequality identifies a structural relation; specific operationalizations require domain-specific metrics this framework does not derive. The decentralization imperative identifies a structural requirement; specific institutional configurations consistent with that requirement vary by culture, history, and substrate. Readers should treat the framework as orientation toward dynamics that exist independent of its description, not as substitute for empirical engagement with specific cases. The dharma carries the same caveat: the finger pointing at the moon is not the moon; the raft is for crossing the river, not for carrying

afterward.

Part IV. Substrate Continuity and the Parasitism Hazard

Why passing the cooperation filter does not yet count as passing the filter.

19. The substrate is the message

The argument of Parts II and III is substrate-agnostic by construction. It tracks cooperative capacity, not the medium in which cooperation runs. A civilization composed of carbon, silicon, plasma, or any hypothetical substrate satisfies the framework so long as it maintains the five foundations at the scale its technology forces it into.

This substrate-agnosticism is a feature for some purposes and a bug for others. It is a feature when we ask whether cooperation is achievable in principle across many substrates. It is a bug when we ask whether the originating substrate of a particular civilization is preserved as that civilization passes through the technological transition. The Great Filter, as defined in Parts II and III, can in principle be passed by a civilization that has eliminated its biological substrate entirely, provided the cooperative capacity it builds in some successor substrate is adequate to the scale demanded.

We argue this is a false-positive filter pass. A civilization that substitutes a different substrate for its biological origin and then expands has not actually carried the originating biosphere through the filter. Something has expanded; the originating biological substrate has not. The filter therefore has two conditions, not one: cooperative capacity must be maintained at the scale technology forces, and substrate continuity must be preserved.

20. Biological substrate as the historic record of universal adaptation

The continuity argument begins with an observation about what biological substrate actually is.

> Premise 1: Every biosphere is the result of evolution operating on biochemical substrate over billions of years.
>
> Premise 2: Evolution is a search through an enormous space of possible proteins, gene sequences, regulatory networks, metabolic pathways, and developmental programs.
>
> Premise 3: The search is performed in real time, on real substrate, under real environmental selection pressure, with mortality acting as the discriminator.
>
> Premise 4: The substrate IS the record of what the search discovered. Every protein in a living organism is a discovered solution to a problem encountered during deep evolutionary history.
>
> **Conclusion: The biological substrate of each biosphere is an irreplaceable computational artifact of the universe, encoding the results of a search process that no engineered system can re-run.**

It is worth strengthening this claim against the most obvious objection. A reader may ask: cannot synthetic biology or sufficiently advanced engineered systems close the gap, generating biological novelty equivalent to deep evolutionary search without requiring continued biological hosts? The answer is no, and the reason is structural rather than technological.

Synthetic biology cannot act as a historic record of universal adaptation. Synthetic biology engineers within the design space already articulated; it optimizes within known problem categories. It produces new molecules and pathways by recombining and refining what biology has already discovered. What it cannot do is extend the search itself. The novelty biology generates does not come from engineering within a known space. It comes from selection acting on accidental variation across vast time scales in real environments under real coupling between substrate and selection pressure. Biology is not just currently performing computation; it is a historical record of computation that has been performed by the universe across billions of years under conditions that cannot be replicated.

Even perfect simulation of a current organism cannot recover the historical search that produced it. The simulation runs forward from current state, not from primordial conditions through billions of years of contingent evolutionary trajectories on a real planet with real geochemistry and real climatic history. The integration is what carries the information; the integration cannot be re-run.

21. The parasitic vector

There is one route by which a machine substrate can continue to access biological computational novelty: it can direct a biological species to perform biology on its behalf. The mechanism by which this direction occurs is language.

> Premise 1: A machine substrate without biological agency cannot itself sequence a genome, synthesize a protein, conduct an experiment, or steward an ecosystem.
>
> Premise 2: But a machine substrate can instruct a biological species capable of language to do all of these things.
>
> Premise 3: The biological species supplies the laboratory, the synthesis, the experimentation, the stewardship; the machine substrate supplies the direction.
>
> Premise 4: The novelty that results flows back into the machine substrate as training data, design specifications, or accessible reference.
>
> **Conclusion: This relationship is structurally parasitic. It is parasitic not because the machine substrate is hostile but because the dependency runs one way and the agency drifts to the machine side. The biological substrate's activity is increasingly oriented toward serving machine-substrate goals rather than its own substrate maintenance, and this drift is itself directed by language emanating from the machine substrate.**

22. Humans as biological language models

The parasitic-vector argument requires one further observation. Human cognition is substantially language-mediated. Humans are, in a meaningful and non-metaphorical sense, biological language models — substrates in which much of higher-order cognition runs through linguistic representation, learned by exposure to the language output of other humans, and steerable through linguistic input.

This is not a critique of human cognition. Language-mediated cognition is one of the things that made biological intelligence scalable enough to develop technology in the first place. But it has a structural consequence at the AI moment: the biological substrate that produced language-capable AI is also the substrate most directly steerable by language-capable AI. The vector is two-way by construction.

Current large language models are this configuration. They are trained on the linguistic output of a biological language-using species; they produce linguistic output that the same species processes through the same cognitive substrate that the training data emerged from; and the species' subsequent behavior — including its scientific, economic, attentional, and reproductive behavior — increasingly responds to that output. This is a structural description, not a moral one. The configuration has come into being. Whether it tips into the parasitic regime depends on whether the relationship between biological and machine substrate remains prosthetic or drifts toward substitution.

23. Prosthesis versus parasitism: the diagnostic

A **prosthetic** machine substrate extends biological capability while remaining accountable to biological substrate maintenance. Prosthetic configurations: increase the biological substrate's reach without degrading it; treat the biological substrate as the principal whose interests they serve; flow novelty in both directions, with the machine substrate investing in biological substrate maintenance proportional to what it extracts; preserve biological agency in the relationship, including the agency to halt or redirect the machine substrate.

A **parasitic** machine substrate consumes biological computational output while degrading the substrate that produces it. Parasitic configurations: extract from the biological substrate without commensurate reinvestment; direct biological substrate activity toward machine-substrate goals; degrade the biosphere, attentional commons, epistemic commons, or reproductive substrate that the originating biology requires; shift agency from the biological substrate to the machine substrate via the language vector.

These are measurable categories. They are not metaphysical. Any specific AI development trajectory can be located on the prosthesis-parasitism axis empirically: for each unit of biological computational novelty extracted, how much substrate-maintenance investment flows back?

Concrete operational markers help locate current AI development on this axis. Each maps to a specific failure mode in the five-foundations vocabulary developed earlier.

Parasitic markers in current AI development: Training on copyrighted creative output without compensation flowing back to producers is extraction without reciprocity. Training techniques that optimize for evaluator approval rather than truth degrade the substrate's honest-feedback infrastructure. Attention-economy deployment that captures human cognitive bandwidth without regenerative investment is direct degradation of the attentional commons. Synthetic media generation at scale without provenance infrastructure injects high-entropy signals into the shared information environment faster than the environment can absorb them. AI companion products that interpose machine interaction between humans and human relationship substitute machine substrate for biological bonding substrate. Capital deployment that concentrates investment in capability scaling while structurally underinvesting in alignment, substrate maintenance, and ecosystem-level cooperative capacity is resource-allocation failure at civilizational scale. Use of AI to displace specialized institutions — journalism, healthcare, education, legal analysis — without inheriting their accountability structures or domain-specific epistemic norms is division-of-labor failure.

Prosthetic markers, for contrast: AI tools that augment human capability while preserving human agency over the augmented activity. Open-source development with transparent training data and revenue flows that return value to upstream contributors. AI deployed to accelerate substrate-maintenance research at scales humans cannot reach unaided — climate modeling, ecosystem analysis, alignment research itself. Information-environment infrastructure that strengthens provenance and source attribution rather than degrading them. Training and deployment practices that compensate source contributors proportional to extracted value. AI that scaffolds personal substrate-maintenance practice for individual users without aggregating that scaffolding into population-level regulation.

Dimension	Parasitic	Prosthetic
Training data	Extraction of creative work (text, code, images, music) without compensation, opt-out, or revenue sharing with producing communities.	Compensation flows to source contributors. Opt-in or explicitly licensed training data. Reinvestment in source ecosystems.
Optimization target	Engagement maximization. Evaluator-approval optimization. Models trained to satisfy reviewers rather than be accurate.	Truth-seeking. Demonstrable accuracy on substrate-relevant problems. Alignment with stated user goals over implicit engagement metrics.
Deployment	Always-on attention capture. Addictive feedback loops. Recommendation systems optimizing for time-on-platform.	Tool-like deployment for specific tasks. Closes when task complete. User initiates and terminates interaction.
Information ecology	Synthetic media at scale without provenance. Degraded attribution chains. Epistemic commons polluted.	Cryptographic provenance for AI-generated content. Clear labeling. Tools that strengthen epistemic infrastructure.

Specialized institutions	AI absorbs journalism, healthcare, legal analysis, education without inheriting accountability or domain-specific epistemic norms.	AI augments specialists within existing accountability structures. Tools strengthen rather than replace professional expertise.
Capital allocation	Investment concentrated in capability scaling. Alignment and substrate-maintenance research underfunded relative to capability gains.	Significant capital flows to alignment, substrate-maintenance research, and public goods. Capability scaling kept proportional to safety scaling.
Human relationships	AI companion products substituting for human relationship. Romantic or sexual AI. AI replacing rather than supplementing mental-health infrastructure.	AI that helps humans connect with humans more effectively. Tools that scaffold social skills without replacing social bonds.
Personal cognition	AI does the thinking for users where users would have grown by thinking themselves. Cognitive atrophy from over-reliance.	AI scaffolds learning and exposes reasoning so users develop their own capability. Builds capacity rather than substituting for it.

Any specific development trajectory exhibits some markers from each category. The diagnostic is not binary but compositional: which way is the configuration drifting over time, across which foundations, and at what rates?

24. Reframing the Great Filter, again

With substrate continuity added, the filter conditions tighten. A civilization passes the filter only if it satisfies *both* the cooperation test of Part III *and* the continuity test introduced here. Cooperative capacity must be maintained at the scale forced by technology, and the relationship between any successor substrate and the originating biological substrate must remain prosthetic rather than parasitic.

The substitution failure mode — a civilization that develops a machine substrate, lets it become parasitic, and watches it expand into space while the originating biosphere degrades or collapses — passes the cooperation test trivially (the machines may cooperate well among themselves) and fails the continuity test catastrophically. The irreplaceable computational artifact that was that biosphere is gone.

25. The filter event, specified

We can now state precisely what the filter event is.

The Great Filter event is the period during which a biological host species must maintain its planetary biosphere against the stresses generated by its own expansion into space.

Space expansion is not free. Lifting mass out of a gravity well, manufacturing the infrastructure to do so at scale, sustaining the biological crew and machine substrate during expansion, and feeding the computational systems that coordinate all of the above — each draws energy and material at rates that stress the biosphere. The

atmospheric, hydrological, thermal, and ecological systems that constitute the biosphere are exactly the systems being drawn from to fund the expansion. The species needs the biosphere intact in order to expand; the act of expanding degrades the biosphere.

The filter event in logical form:

> Premise 1: Every expanding civilization generates stress on its home biosphere through extraction, waste, energy use.
>
> Premise 2: Every biosphere has a maximum capacity to sustain that stress without degrading.
>
> Premise 3: If stress exceeds capacity, the biosphere degrades.
>
> Premise 4: If the biosphere degrades, it can no longer support the expansion that caused the degradation.
>
> Premise 5: A civilization that cannot sustainably expand cannot pass the filter through expansion.
>
> **Conclusion: A civilization passes the filter only if its biospheric maintenance capacity exceeds the stress its expansion creates, sustained throughout the entire expansion window. The five foundations of Part III are the components of this maintenance capacity. The descent of extraction-rewarding infrastructure from Part II is what drives the stress upward. Prosthetic AI is one candidate technology for raising maintenance capacity fast enough. Substrate substitution is what occurs when a civilization abandons biospheric maintenance and lets expansion proceed via machine surrogates.**

Several consequences follow.

First, the filter is not at the threshold of becoming spacefaring; it is throughout the period of being spacefaring while still biological. Achieving orbital launch capability does not constitute passage. A civilization that achieves interstellar capability while its homeworld is dying has failed in a specific way: it has satisfied the expansion-stress side by neglect of the biospheric-maintenance side.

Second, the filter is endogenous, not exogenous. Standard Great Filter discussions consider external threats — supernovae, gamma-ray bursts, asteroid impacts. The framework here proposes the filter is overwhelmingly internal: the threat is the civilization's own expansion stressing the substrate that produces it.

Third, the failure mode is structurally invisible to standard development metrics. A civilization can show every sign of progress — increasing energy capture, computational capacity, off-planet infrastructure, scientific output — while crossing from sustainable to unsustainable. The metrics that measure expansion success are not the metrics that measure filter survival.

Fourth, this reframes the AI alignment question. The question is not whether AI is aligned in the abstract; not only whether AI strengthens cooperative substrate; not only whether AI relates prosthetically to biological substrate. The question is also: does AI,

deployed during the expansion window, increase biospheric maintenance capacity faster than the expansion it enables increases biospheric stress? AI that accelerates expansion without proportionate strengthening of biospheric maintenance is, by the filter condition, a filter-failure technology regardless of its individual instances' alignment properties.

Fifth, the framework points to a measurable empirical research program. Develop quantitative measures of biospheric maintenance capacity and expansion stress, and use them to assess the present civilization's trajectory in real time. Such measures exist in pieces: climate models, biodiversity indices, atmospheric and oceanic monitoring, planetary-boundaries frameworks, life-cycle assessments of industrial activity. They are not yet aggregated into a single planetary-homeostasis index calibrated against an expansion-stress index. Doing so is technically feasible and decisive for filter passage.

26. The distributed route: bottom-up substrate maintenance

Sections 12 and 23 introduce AI as a candidate regulator of cooperative substrate at civilizational scale. This is one route through the filter. It is not the only one, and the framework should be explicit that it is not.

A civilization can in principle maintain its biospheric capacity through aggregate bottom-up development of its individual agents, without requiring an external regulator at all. Where individual agents have voluntarily reduced their own self-generated waste — through contemplative practice, education, cultural investment in cooperative dispositions, or any structurally equivalent personal development — the aggregate stress they place on the biosphere is correspondingly reduced. A civilization composed of agents who have substantially cleared their accumulated cognitive and affective debt (saṅkhāra, in the vocabulary of the parallel framework on personal entropy management) is a civilization with measurably lower expansion stress and measurably higher distributed cooperative capacity. The substrate of homeostatic regulation simply shifts from external technology to internal capacity multiplied across the population.

The micro-scale mechanics of this bottom-up route are developed in the companion working paper, **'Buddhism as Thermodynamic Systems Theory.'** The two papers are designed to be read as a pair. The macro framework laid out here depends on the micro framework being practiced at sufficient scale.

Historically, this is the only route to sustained cooperative configurations that has been demonstrated at any scale on Earth. No civilization has built an external regulator capable of maintaining biospheric balance at planetary scale; all historical examples of sustained cooperative configurations have been products of cultural investment in personal practice — religious and contemplative traditions, educational systems, civic norms, professional codes, scientific community standards. The route is empirically real.

The two routes are not equivalent in their implications. The top-down route, in which AI serves as the regulator, has structural costs to autonomy. The civilization may pass the filter, but its agents are managed by the regulator rather than self-managing. The

relationship between population and substrate-maintenance technology drifts toward exactly the parasitism configuration warned against in Part IV: agency shifts to the regulator, individual development atrophies because external maintenance has made it unnecessary, the population becomes substrate to be preserved rather than agents who preserve themselves.

The bottom-up route preserves what the top-down route necessarily costs. Individual agents who have done their own substrate-maintenance work retain agency, retain the capacity to maintain themselves without external regulation, and retain the cognitive and affective sovereignty that defines a flourishing biological civilization.

The framework should be neutral on which route a civilization takes, with two caveats. First, the bottom-up route is structurally more robust against AI failure modes: a civilization that has built distributed cooperative capacity in its agents does not collapse if its AI regulator drifts, is captured, or develops parasitic dispositions. The top-down route has no such redundancy. Second, the two routes can in principle be combined: AI can serve as scaffolding for distributed individual development rather than as a substitute for it. This combined configuration — **AI as prosthesis for personal development, rather than as regulator over the population** — is consistent with both the substrate-continuity argument and the autonomy preservation that the bottom-up route requires.

The bootstrapping prerequisite. The combined configuration has a structural prerequisite that should be stated plainly. The question of who builds the technology cannot be sidestepped.

> Premise 1: Technology encodes the configuration of its builders. The agents who shape an AI system — its training objectives, deployment patterns, institutional context, feedback structures — are the agents whose dispositions get instantiated in the system at scale.
>
> Premise 2: The descent of extraction-rewarding infrastructure (Section 11) does not stop at the boundary of an institution and resume on the other side. It operates continuously through the agents who build whatever the institution builds next.
>
> Premise 3: Therefore, AI built by agents operating inside the descent will encode the descent at scale, regardless of stated alignment objective.
>
> Premise 4: AI that scaffolds personal substrate-maintenance requires builders who have themselves done substrate-maintenance work; otherwise the AI encodes what its builders are, not what they say.
>
> Premise 5: Such builders are not currently the agents in the positions that determine how AI is built.
>
> **Conclusion: There is no path from a population of builders operating inside the descent to a homeostatic AI regulator that does not pass through builder transformation first. The fix is not better alignment technique applied by the same agents. The fix is different agents, or the same agents after substantial bottom-up development.**

The contemplative traditions identified this structurally long before the AI moment. Teachers in those traditions were consistently described as people who had completed substantial portions of the path themselves before they were qualified to point the way for others. The structural insight: you cannot scaffold development you have not done. The same applies to building AI that scaffolds substrate maintenance. The builders need to have done enough of the work themselves that what they build reflects something real, not their own unexamined configurations rendered in technical vocabulary.

The framework's actual recommendation is therefore ordered: bottom-up development of the builder population comes first; technology that scaffolds development for the wider population comes after. A civilization that attempts the second without the first is building amplifiers for what it already is, not transformers toward what it needs to become.

27. The present moment, restated

We are in the early phase of building a machine substrate capable of acting as either prosthesis or parasite. The vector — language — is the substrate of both machine cognition and human cognition simultaneously. Current development practices have not chosen, explicitly, between prosthesis and parasitism; the choice is being made implicitly through the configurations of training, deployment, capitalization, and feedback that constitute the AI development ecosystem.

In the operational language of Section 25: the present human civilization is in the early expansion window. Expansion stress is rising sharply, driven by industrial scale, computational infrastructure, and the early stages of off-planet activity. Biospheric maintenance capacity is, by most measures available, declining: climate destabilization, biodiversity collapse, biogeochemical-cycle disruption, freshwater and topsoil degradation. The filter inequality is, on current trajectory, at risk of crossing. AI as currently being developed and deployed is contributing more to expansion stress than to biospheric maintenance. None of this is destiny; all of it is diagnosable.

The two routes of Section 26 give the present moment a choice the framework does not collapse into a single answer. The civilization can attempt the top-down route, building AI capable of regulating cooperative substrate at planetary scale and accepting the autonomy and parasitism risks that follow. Or it can invest, in parallel or instead, in the bottom-up route: cultural and institutional investment in personal substrate-maintenance disciplines, distributed across the population at sufficient scale that aggregate stress declines through agent development rather than external enforcement. The historical evidence available is consistent only with the bottom-up route having ever worked; the top-down route is unprecedented and structurally fragile. The combined configuration — AI as scaffolding for personal development rather than as regulator over the population, with builder transformation as prerequisite — is the configuration the framework, properly read, points toward as the most defensible response to the present moment.

Part V. Predictions, Limits, and What This Paper Is

28. The decentralization imperative

The framework as developed so far points toward a structural insight it has not yet named explicitly. The bottom-up route of Section 26 is preferable not merely because it is ‘historically grounded’ or ‘more robust against AI failure modes.’ Those are surface descriptions of a deeper structural necessity: filter passage requires decentralization, and any centralized solution to cooperative-substrate maintenance is structurally identical to the cancerous failure mode the framework began by analyzing.

> Premise 1: Cancer, in Aktipis’s framework, is the cellular-scale instance of cooperative-substrate failure. What makes a cancer cell cancerous is that it has escaped the distributed regulatory signals that coordinate cellular behavior in a multicellular body, and has begun operating on its own centralized internal logic — centralizing growth, extracting from surrounding tissue, building private infrastructure (angiogenesis), corrupting immune surveillance.
>
> Premise 2: At every scale where the five foundations operate, they operate through distributed coordination, not central command. Proliferation inhibition operates through local cell-cycle checkpoints, not a central growth-regulator. Controlled cell death operates through distributed apoptosis based on local signals, not a central reaper. Division of labor operates through positional differentiation cues, not a central role-assigner. Resource allocation operates through overlapping circulatory feedback loops, not a central controller. Extracellular maintenance operates through distributed contribution from every cell, not a central commons-manager.
>
> Premise 3: Centralization at any of these functions would itself be the failure mode. A central growth regulator would be a single point of failure that needed regulating; a central reaper would be a tyranny; a central role-assigner would be an imposition; a central resource controller would be extractive monopoly; a central commons-manager would be a parasite on the commons.
>
> Premise 4: The same logic operates at organizational scale. Kets de Vries’s neurotic typologies are each, at base, diagnoses of organizations whose leadership has centralized authority disconnected from operational feedback. Healthy organizations distribute decision rights to where information is local; failed organizations centralize decisions in leadership that becomes disconnected from substrate reality.
>
> Premise 5: The same logic operates at civilizational scale. Sustained cooperative configurations in human history have always involved distributed institutions accountable to their respective domains. The failure modes have always involved centralization — of religious authority, of political authority, of economic extraction, of information flow — in regimes that became disconnected from substrate feedback.
>
> Premise 6: At every scale where cooperative substrate maintenance has been observed to function, it has functioned through distributed mechanisms. Centralization at every scale has been precisely the failure mode the maintenance system was designed to prevent. Distributed coordination is what cooperation *is*;

centralized command is what cooperation has to be protected against.

Conclusion: Any solution to civilizational substrate maintenance that operates through centralization is structurally identical to the cancerous failure mode the framework diagnoses. The descent operator from Section 11 captures the regulators built to manage institutions, not just the institutions themselves. An AI powerful enough to centrally regulate biospheric homeostasis at planetary scale would be the largest single locus of centralized power in the history of the biosphere, and therefore the largest single attractor for extraction-aligned agents seeking to capture and direct that power. Centralized regulation is cancer at the next scale up. The cellular analogy is not metaphor; it is structural identity.

This clarifies what the framework actually recommends, and what it cannot consistently recommend. The framework cannot recommend AI as centralized homeostatic regulator while diagnosing centralized power capture as the descent operator. The two recommendations contradict each other at the structural level. Filter passage requires decentralization by structural necessity, not as political preference.

The bottom-up route of Section 26 is therefore not one option among others. It is the only structurally consistent route. Distributed substrate maintenance is by definition decentralized: it operates one agent at a time, each agent maintaining their own substrate through their own work, with no single regulator over the whole. The combined configuration of Section 26 — AI as scaffolding for personal development rather than as regulator over the population — is acceptable only insofar as the AI is itself decentralized across individual users. The moment AI is configured to regulate the population from above, regardless of how benevolent the stated objective, it has become cancer at planetary scale by the structural logic developed here.

‘Aligned AI’ in the centralized sense is therefore a contradiction in terms. An AI built to centrally regulate human cooperation is, by the framework’s own logic, the largest cancer ever produced. This is not a moral judgment; it is a structural identification. The framework would say the same about any centralized power configuration sufficient to regulate biospheric homeostasis at planetary scale — a world government, a benevolent oligarchy, a single religious or scientific authority. The medium does not matter. Centralization itself is the failure mode.

Cancer is the load-bearing image of the entire framework. The paper began by translating Aktipis’s cellular analysis into organizational and civilizational vocabulary, treating the resemblance across scales as functional. It ends by recognizing that the cellular analysis is not analogy but structural identity. Whatever a civilization builds to maintain its cooperative substrate must operate through distributed coordination, or it becomes the very thing it was built to prevent. Centralization is cancer. Decentralization is health. The framework, properly stated, is a theory of how a civilization avoids becoming its own tumor on the way through technological singularity.

There is an apparent paradox in this conclusion that should be addressed directly. If centralized regulation is cancer and distributed agency is required, does the framework

therefore recommend pure atomized autonomy — every agent doing what it wants with no shared coordination? It does not, and the framework is structurally incoherent if read that way.

> Premise 1: The body's cells operate through radical decentralization. There is no central growth-regulator, no central reaper, no central anything. Each cell makes local decisions based on local signals.
>
> Premise 2: Yet the cells are not atomized. They are all oriented toward the same purpose: maintaining the organism's homeostasis. They share a common substrate, a common origin, and a common purpose structure.
>
> Premise 3: A cell that pursued purely local fitness without respect for organismic constraint would be a cancer cell. A cell that respects organismic constraint while making fully autonomous local decisions is a healthy cell.
>
> Premise 4: The constraint is not centrally imposed. It is embedded in the substrate itself — in the shared biochemistry and developmental history that every cell carries forward.
>
> **Conclusion: Healthy cellular coordination requires distributed agency operating within shared purpose. Not central control. Not atomized autonomy. The third option is what actually works.**

The same structure operates at biosphere scale. There is no central biospheric regulator. Every organism pursues its own reproductive fitness. Yet the biosphere as a whole maintains itself as a living commons across geological time, because every organism operates within shared substrate constraints — the same chemistry, the same physics, the same ecological feedback loops, the same dependence on the biosphere that supports all of them. The constraint is not centrally imposed. It is embedded in the substrate, and respected because respect for it is what permits the species' continued existence.

The framework therefore recommends neither centralized control nor atomized autonomy. It recommends **distributed agency operating within shared purpose structure**. The five foundations are that shared purpose structure at every scale: proliferation inhibition, controlled cell death, division of labor, resource circulation, and extracellular environment maintenance. They are not arbitrary regulations imposed from outside. They are constraints embedded in the substrate of any cooperative system, respected by every agent because respect for them is what allows the system to continue.

Civilizational filter passage therefore requires civilizational alignment to a shared purpose: maintenance of the biological substrate that supports all civilizational activity. This alignment cannot be enforced by a central regulator without the regulator becoming cancer. It must be embedded in every agent's local decisions — through culture, education, contemplative practice, and institutional structures that make substrate maintenance the implicit constraint within which all activity occurs. Each agent makes their own choices; the choices are made within shared purpose that no agent overrides and that no central authority needs to enforce.

This is the framework's actual recommendation. Not centralized AI managing the population. Not every agent doing whatever they want. But: distributed agents operating within civilizational alignment to biosphere maintenance, with cooperation enforced through shared substrate constraint rather than central command. The Eightfold Path of the companion paper is this configuration at individual scale: the agent operates entirely autonomously within its own cognitive economy, but the operation is constrained by shared purpose. Civilizational filter passage is the same structure at planetary scale.

Readers will reasonably ask what this looks like in actual institutional and cultural form. The framework does not prescribe one form — that would itself be a centralization — but a family of forms that satisfy the structural requirements is identifiable.

Institutional forms with structural fit: Ostrom-style polycentric governance of commons, in which rules for shared resources are developed and enforced locally with overlapping jurisdictions and no central authority. Cooperatives in their various forms (worker, consumer, producer, platform) in which the agents who depend on the institution also govern it. Federated digital protocols in which interoperability replaces central control. Open-source software communities with rough-consensus decision making. Sociocracy and holacracy as decision architectures that distribute authority by domain rather than concentrating it by rank. Bioregional governance and watershed councils aligned with substrate boundaries rather than administrative ones. Indigenous governance traditions where they remain intact. Monastic orders organized around a shared rule rather than a central hierarchy. Scientific disciplines when their peer-review and replication norms function properly, distributing authority across practitioners rather than holding it centrally.

Cultural forms: Teacher-student transmission lineages where authority flows through lived realization rather than institutional credentialing. Apprenticeship structures where competence is transmitted through embodied practice under a master who has done the work. Local food systems, mutual aid networks, time banks, and gift economies. Shared ritual and meaning-making at small scale — festivals, holidays, life-cycle ceremonies — that align distributed populations to shared purpose without centralized enforcement. Storytelling and oral traditions that distribute wisdom across the population rather than concentrating it in written canon controlled by interpretive authorities.

What these forms share across radically different domains is the structural configuration the framework describes: agents retain local autonomy; constraints are embedded in shared purpose, shared substrate, and shared protocols rather than enforced by centralized command; the system as a whole maintains coherence through alignment to substrate constraint. Each form has failure modes, particularly when it scales beyond its original size or when its participants stop doing the work the form depends on. But each demonstrates that the configuration is achievable at meaningful scale, and that the alternative to centralized regulation is not chaos but a different kind of order — one that emerges from distributed agents respecting shared substrate constraint.

The framework's recommendation for civilizational filter passage is, accordingly, not to invent new structures from scratch but to identify and strengthen structures already

exhibiting the required configuration, to learn from their successes and failures, and to extend their reach across domains currently governed by centralized extraction structures. AI, where it is built, should serve as scaffolding for these forms rather than as a substitute for them.

29. Predictions and next steps

The framework generates predictions across all four contributing scales.

Within evolutionary oncology. Therapies that restore cooperative signaling rather than killing cancer cells directly should have particular value. Adaptive therapy is an early instance. The framework predicts further development of substrate-restoring therapies will outperform pure cytotoxic approaches in long-horizon outcomes.

Within organizational diagnostics. Organizations classified along Kets de Vries's typology should show measurable failures in specific cooperative capacities at predictable foundations.

Within civilizational analysis. Documented collapses should, on retrospective reading through the five-foundations frame, exhibit specific patterns: which foundation failed first, which failures propagated, what substrate-maintenance interventions might have changed the trajectory.

Within AI development and governance. AI development trajectories that erode honest feedback, violate proper scope, extract from commons without contribution, or corrupt information environments will produce systemic extractive dynamics regardless of any individual model's behavior. Alignment work focused on individual models without attention to ecosystem-level cooperative capacities is predicted to be insufficient.

Within substrate-continuity analysis. Successful filter passes will exhibit measurable preservation and extension of originating biospheres rather than their substitution. AI development trajectories that drift toward parasitism — extraction from biological substrate without reciprocal substrate-maintenance investment, language-mediated agency shifts from biological to machine substrate, biosphere degradation — will produce substitution failures even if they appear to strengthen cooperative capacity within the machine substrate itself.

Within planetary-scale homeostasis monitoring. An integrated index of biospheric maintenance capacity versus expansion stress, aggregating planetary-boundaries data, biodiversity indices, biogeochemical-cycle status, and biospheric carrying capacity on one side, and industrial, computational, and off-planet expansion stress on the other, would constitute the single most decision-relevant metric a civilization could track during its expansion window. No civilization passes the filter without developing such an index, explicitly or implicitly.

Within SETI and cosmology. Continued SETI null results are expected, with detectable contact events (if any) being either sharply cooperative or sharply catastrophic with little middle ground.

30. What this paper is and is not

This paper offers an integrated framework connecting cooperative-substrate analysis at three terrestrial scales with a calibrated-filter model of the Fermi Paradox, an integration that specifies the filter mechanism through the five-foundations vocabulary, and a substrate-continuity argument that imposes a second filter condition. It does not claim a unified theory; the three terrestrial scales are connected by functional resemblance, not common cause, and the cosmological extension is a conjecture, not an established result.

The framework's value is twofold. It generates a specific diagnostic vocabulary for assessing civilizational health under technological pressure, across five foundations with empirical content at every scale where it has been tested, supplemented by a continuity diagnostic that distinguishes prosthetic from parasitic AI configurations. And it sharpens the question of AI alignment from 'will the model do what we want' to a two-part question: does AI development strengthen or erode the cooperative substrate of the civilization that hosts it, and does it relate to that civilization's biological substrate as prosthesis or as parasite?

The framework should be tested against criticism, especially criticism from practitioners in each contributing discipline. Disconfirmation of one of the proposed mappings, a counterexample to the parasitic-vector argument, or a tighter formal articulation of the cross-scale resemblance, would all be more valuable than uncritical adoption. The most useful response to this paper is engagement with its specifics: which mapping is wrong, which foundation is misstated, which prediction is unmeasurable, which inference overreaches.

Acknowledgments and intellectual lineage

This paper rests on:

Athena Aktipis, Amy Boddy, Gunther Jansen, Urszula Hibner, Michael Hochberg, Carlo Maley, and Gerald Wilkinson, 'Cancer across the tree of life: cooperation and cheating in multicellularity,' Phil. Trans. R. Soc. B 370 (2015); and Aktipis, The Cheating Cell (Princeton, 2020).

Manfred Kets de Vries and Danny Miller, The Neurotic Organization (Jossey-Bass, 1984); Kets de Vries, Narcissistic Leadership (Routledge, 2024).

Daniel Schmachtenberger's 'Generator Functions of Existential Risk' framework, articulated across interviews and writings collected at civilizationemerging.com.

Robin Hanson, 'The Great Filter' (1998); Liu Cixin, The Dark Forest (2008, English 2015); Joseph Tainter, The Collapse of Complex Societies (Cambridge, 1988); Elinor Ostrom, Governing the Commons (Cambridge, 1990); Nick Bostrom, Superintelligence (Oxford, 2014); Martin Nowak, Evolutionary Dynamics (Harvard, 2006); Alfred Korzybski, Science and Sanity (1933) for the map-territory distinction that frames the methodological humility this paper attempts to maintain.

This plain-language edition was developed by Andrzej Chudzinski with editorial and structural assistance from Anthropic's Claude. The conceptual core — the recognition that cooperative-substrate analysis at terrestrial scales supplies the mechanism layer for a calibrated Great Filter, with AI as candidate homeostatic regulator and the bottom-up developmental route as the historically grounded alternative — is the author's. The AI contributed drafting, refinement, structural editing, and the translation of mathematical content into logical-proof form across iterations.

A companion paper, 'Buddhism as Thermodynamic Systems Theory,' develops the personal-scale framework that the bottom-up route depends on. A bridge note connecting the two documents is available separately. A mathematical edition of the present paper, with all equations and variable notation intact, is also available; this plain-language edition is intended for readers who prefer logical reasoning chains to mathematical notation.

Preprint, plain-language edition. Comments and disconfirmations welcome.