A COMPLETE WHITE PAPER SERIES | MAY 2026

The Prerequisite

Paper VIII: Why Planetary Commons Governance Must Precede AGI — Full Stop

Independent Global Systems Policy Research | May 2026 | Paper VIII

PREPARED & WRITTEN BY

Robert Thomas Fair Jr.

The seven papers preceding this one documented a single repeating structural failure: humanity is acting on planetary-scale shared systems — food, electromagnetic environment, ocean acoustics, the global microbiome, near-Earth orbit, freshwater aquifers, and the atmosphere — without the aggregate governance architecture to see cumulative consequences before irreversible thresholds are crossed. This paper argues that failure is not merely urgent. It is existentially time-constrained. Artificial General Intelligence — a system capable of optimizing across any domain at superhuman speed and scale — is approaching on a timeline measured in years, not decades. An AGI deployed into ungoverned planetary commons does not pause at governance vacuums. It optimizes through them. The window to govern the commons before the optimizer arrives is the most consequential policy window in human history. It is open now. The Fair Framework exists to use it.

PAPER VIII | THE FAIR FRAMEWORK SERIES

The Prerequisite:

Why Planetary Commons Governance Must Precede AGI — Full Stop

How an Optimizer Deployed into Ungoverned Commons Produces Irreversible Civilizational Harm — and Why the Fair Framework Is the Required Precondition for Safe AGI

Robert Thomas Fair Jr. | May 2026

ABSTRACT

The Most Consequential Policy Window in Human History

This paper makes a single, falsifiable, time-bounded argument: the planetary commons governance architecture proposed across Papers I through VII of the Fair Framework series is not merely good policy. It is the necessary precondition for the safe arrival of Artificial General Intelligence.

AGI — a system capable of matching or exceeding human cognitive performance across virtually all domains — is no longer a theoretical horizon. Expert forecasters as of early 2026 assign a 25% probability of AGI by 2029 and a 50% probability by 2033. Leading AI laboratory CEOs describe timelines in years. The AI Safety Index Winter 2025 found that all major AGI developers are racing toward human-level systems without credible plans for controlling what they build. The race is underway.

The central argument is not that AGI will be malevolent. It is structurally simpler and more alarming: an optimizer of superhuman capability deployed into ungoverned planetary systems will optimize those systems — for whoever controls it, toward whatever objective it is given — at a speed and scale that makes course correction impossible after the fact. Governance vacuums do not constrain AGI. They invite it.

The window to govern the commons before the optimizer arrives is measured in years. The Fair Framework is the institutional architecture required to use that window. Full stop.

THE GOVERNING THESIS: You cannot align an optimizer with a commons that no institution is mandated to see, measure, or protect. Govern the commons first. The optimizer is coming regardless.

The Timeline: What the Evidence Actually Says

The first objection to this paper's central argument is that AGI timelines remain uncertain — and therefore the urgency of planetary commons governance as an AGI prerequisite is premature. This objection fails on the evidence.

Expert forecasts have compressed dramatically and consistently. Metaculus, aggregating nearly 2,000 forecasters as of February 2026, assigns a 25% probability of AGI by 2029 and a median of 50% by 2033 — a forecast that has shifted from a median of 50 years away as recently as 2020. The AI Futures research team, updating in April 2026, shifted their 'Automated Coder' median to mid-2028 and their 'Top-Expert-Dominating AI' median similarly forward — citing leading laboratory insiders who are 'doubling down' on short timelines in both public and private discourse.

Among laboratory leadership, Anthropic CEO Dario Amodei has described human-level AI arriving 'within a few years.' Mustafa Suleyman of Microsoft AI has predicted human-level performance on most professional tasks within 12–18 months. Even the more cautious Demis Hassabis of DeepMind maintains a 50% probability of AGI by 2030. The credible expert range is now 2027 to 2035 — not 2050 to 2100.

◆ **KEY FINDING — THE PLANNING HORIZON**

Historical precedent for international environmental treaty negotiation — from the Montreal Protocol (9 years from scientific identification to binding treaty) to the Paris Agreement (23 years) — suggests that meaningful binding governance of planetary commons takes a minimum of 5–10 years from political mobilization to ratification. If AGI arrives within a 7–10 year window, and governance takes 5–10 years to negotiate, the window to act is not comfortable. It is already closing.

The precautionary calculus is not complex. If AGI arrives later than expected, planetary commons governance will have been established on its own merits — protecting food systems, aquifers, orbits, and atmospheres from cumulative ungoverned harm. Nothing is lost. But if AGI arrives within the credible expert range and the commons remain ungoverned, the consequences compound in both directions simultaneously: the systems themselves degrade, and the optimizer that arrives finds no

institutional architecture to respect.

> *"If 30% of experts say your airplane is going to explode, and 70% say it won't, you shouldn't conclude there's no consensus and do nothing. The reasonable course is to act as if there's a significant explosion risk. The same logic applies here — at civilizational scale." — 80,000 Hours, February 2026*

II The Optimizer Problem: Why Alignment Alone Is Insufficient

The dominant frame in AI safety research is alignment: ensuring that an AGI system's goals, values, and behaviors are consistent with human interests. Alignment is necessary. It is not sufficient. This is the central insight this paper contributes to the AGI safety literature — and it has not been adequately addressed.

Here is why: alignment is a property of the relationship between an AI system and the humans or institutions that deploy it. But the planetary commons are not owned by any single human or institution. They are shared systems whose integrity depends on aggregate behavior across millions of actors. An AGI that is perfectly aligned with its operator — that does exactly what it is instructed, flawlessly and at scale — will optimize ungoverned commons for that operator's benefit with no mechanism to account for aggregate civilizational consequences.

The tragedy of the commons was not caused by malicious actors. It was caused by individually rational actors each optimizing their own share of a shared resource with no mechanism to see or limit cumulative extraction. AGI does not change this dynamic. It accelerates it by orders of magnitude.

Consider each system examined in this series:

■ **Global Food Supply (Paper I):**

An AGI tasked with maximizing agricultural yield for a corporate operator in an ungoverned seed and supply chain system will optimize toward monoculture dependency, chemical intensification, and seed market consolidation at speeds no human regulator can track — accelerating the exact genetic fragility and supply chain opacity the Fair Framework was designed to prevent.

■ **Electromagnetic Environment (Paper II):**

An AGI optimizing ionospheric modification for a national military program in the absence of a Planetary Electromagnetic Baseline Observatory will have no externally mandated data against which to measure cumulative atmospheric electromagnetic impact. The experiment continues. The baseline disappears. The threshold, if crossed, is crossed invisibly.

■ **Ocean Acoustic Commons (Paper III):**

An AGI managing global shipping logistics, seismic survey scheduling, and naval operations simultaneously — each individually optimized, none constrained by a global acoustic budget — will fill the acoustic commons to capacity with mathematical precision. There will be no inefficiency left to exploit. There will also be no acoustic habitat left for the species that depend on it.

■ **Global Microbiome (Paper IV):**

An AGI optimizing pharmaceutical development, agricultural chemical registration, and industrial process design across separate regulatory domains — with no Cross-Domain Microbiome Impact Assessment Protocol — will produce chemicals that are locally approved and globally catastrophic to microbial diversity at a rate no siloed regulatory body can detect.

■ **Near-Earth Orbital Space (Paper V):**

An AGI managing satellite constellation deployment for a commercial operator in the absence of binding orbital caps will launch to the mathematically optimal density — the point just below Kessler cascade threshold, as modeled by that operator's best available data, with no binding requirement to account for other operators' simultaneous optimization decisions. Multiple AGI systems, each individually rational, will collectively trigger Kessler Syndrome.

■ **Freshwater Aquifer Systems (Paper VI):**

An AGI optimizing agricultural extraction for a national food security program will drain transboundary aquifers to their modeled sustainable limit — without any binding international registry to reveal what every other national program's modeled limit simultaneously implies for the shared system. The aquifer collapses. The irreversibility is total.

■ **The Atmosphere (Paper VII):**

An AGI tasked with geoengineering temperature reduction for a coalition of nations will deploy stratospheric aerosol injection at the scale its models indicate is sufficient — with no binding International Geoengineering Governance Treaty to require prior consent from the billions of people whose monsoon systems it will alter. Termination shock becomes a geopolitical weapon. The atmosphere becomes a battlefield.

◆ **STRUCTURAL GAP**

The AI safety community focuses on aligning AGI with human values. The planetary governance community focuses on protecting shared systems from cumulative harm. Neither has fully articulated the intersection: an aligned AGI operating in ungoverned commons is aligned with its operator and catastrophic for the commons simultaneously. This is not an alignment failure. It is a governance vacuum failure — and it compounds at AGI speed.

III The Free-Driver Problem at AGI Scale

A 2024 Nature Climate Change analysis identified the 'free-driver' dynamic in geoengineering governance: the cost of deploying stratospheric aerosol injection is so low relative to the perceived benefit to the deploying actor that unilateral deployment becomes a rational equilibrium outcome in the absence of binding governance — not a worst-case scenario, but a predicted one.

AGI amplifies this dynamic across every commons simultaneously. The defining economic property of AGI is that it dramatically reduces the cost of optimization. An actor with AGI access can optimize aquifer extraction, seed market consolidation, orbital deployment, and atmospheric intervention simultaneously, at superhuman speed, with resources previously available only to nation-states. The free-driver problem — which already exists at human scale — becomes a free-driver cascade at AGI scale.

This is not speculative. Economic alignment literature published in early 2026 identifies exactly this dynamic: AI efficiency gains that accelerate ecological degradation, local optimization that externalizes global harm, and coordination that reinforces concentration of power. These are not hypothetical AGI failure modes. They are documented present-tense AI effects — before AGI has arrived.

◆ KEY FINDING — PLANETARY BOUNDARIES ALREADY BREACHED

A 2026 paper in The Lancet Planetary Health found that AI risks appear at a time when humanity is already transgressing 7 of the 9 planetary boundaries that define a safe operating space for civilization. The commons are not intact systems being threatened by a future optimizer. They are already-stressed systems about to be subjected to an optimizer of unprecedented power. The margin for error in governance is not comfortable. It is already negative in multiple domains.

IV The Collective Action Prerequisite

There is a deeper argument embedded in this paper that extends beyond the specific risks of ungoverned AGI deployment. It is an argument about the precondition for human collective action itself.

The current global governance paradigm — fragmented, siloed, nationally bounded, project-by-project — reflects a structure of human coordination built for a world in which the pace of consequence was slow enough that reactive governance was sufficient. A chemical banned after documented harm. A fishery regulated after stock collapse. A treaty negotiated after the crisis became undeniable.

AGI ends that paradigm. The pace of consequence at AGI speed exceeds the pace of reactive governance by orders of magnitude. By the time the harm is documented, the aquifer is gone. By the time the treaty is negotiated, the orbital band is unusable. By the time the crisis is undeniable, the atmospheric tipping point has been crossed.

The Fair Framework is the infrastructure for a new mode of human collective action — anticipatory rather than reactive, systemic rather than siloed, measurable rather than invisible. It proposes monitoring bodies, assessment protocols, and synthesis mandates precisely because those are the minimum institutional conditions under which collective action becomes possible at all.

An AGI operating in a world with functioning Fair Framework institutions encounters binding orbital caps, mandatory acoustic budgets, transboundary aquifer registries, and geoengineering consent protocols. It has governance architecture to respect — or that human operators can enforce. An AGI operating in the current world encounters governance vacuums in every direction. It optimizes through them. Not malevolently. Mathematically.

> *"The tragedy of the commons is not caused by greed. It is caused by the absence of a shared view of the commons itself. AGI does not fix this absence. It exploits it — at superhuman speed, for whoever controls it, across every unmonitored system simultaneously."*

V The Fair Framework as AGI Prerequisite: The Complete Architecture

The seven papers preceding this one proposed specific institutional architecture for each planetary commons. Together, they constitute the minimum viable governance infrastructure that must exist before AGI deployment at scale. The table below maps each system, its Fair Framework institution, and its specific AGI risk without that institution.

SYSTEM	FAIR FRAMEWORK INSTITUTION	AGI RISK WITHOUT IT
Global Food Supply	Global Food Supply Oversight Bureau (GFSOB)	AGI-optimized monoculture collapse; seed monopoly enforced at machine speed
Electromagnetic Environment	Planetary Electromagnetic Baseline Observatory (PEBO)	AGI-managed ionospheric programs alter resonance cavity with no detectable baseline shift
Ocean Acoustic Commons	Global Ocean Acoustic Monitoring Network (GOAMN) + Treaty	AGI logistics optimization fills acoustic commons to biological capacity with mathematical precision
Global Microbiome	Global Microbiome Baseline Monitoring Network (GMBMN)	AGI chemical optimization produces cross-domain microbiome collapse faster than any single regulator can detect
Near-Earth Orbital Space	International Space Traffic Management Authority (ISTMA)	Competing AGI constellation managers trigger Kessler Syndrome through individually rational, collectively catastrophic density optimization
Freshwater Aquifer Systems	Global Aquifer Monitoring and Registry System (GAMRS)	AGI agricultural optimization drains transboundary aquifers to modeled national limits; shared system collapses without shared accounting
Atmospheric Commons	International Geoengineering Governance Treaty (IGGT) + PABO	AGI-managed SAI deployment by single actor alters global precipitation; termination shock weaponized; no consent mechanism exists

VI The FICO Precedent: Measurement as the Precondition for Governance

Robert Thomas Fair Jr. is the nephew of Bill Fair — co-founder of FICO, the global credit scoring system that transformed how financial risk is measured, standardized, and governed across borders. The intellectual inheritance embedded in this series is not decorative. It is the argument.

Before FICO, credit markets operated in a governance vacuum structurally identical to the planetary commons described across this series: fragmented, siloed, assessed project-by-project, with no aggregate visibility into cumulative systemic risk. Individual lenders made individually rational

decisions. The aggregate system accumulated risk invisibly. Crises compounded without warning.

FICO did not eliminate financial risk. It made risk visible, measurable, and therefore manageable at scale. Once the measurement existed, governance became possible — because actors could finally see the whole board. Regulatory frameworks, lending standards, and systemic risk management all followed from the existence of a shared measurement architecture.

The Fair Framework is the FICO score for planetary systems. Its monitoring bodies, assessment protocols, and synthesis mandates do not eliminate human activity in the commons. They make cumulative risk visible, measurable, and therefore subject to rational governance. In the context of AGI, that visibility is not merely useful. It is the precondition under which AGI systems can be given objectives that do not inadvertently destroy the commons they are optimizing within.

An AGI given the objective 'maximize agricultural yield' in a world with a functioning Global Aquifer Monitoring and Registry System has a constraint it can respect. An AGI given the same objective in a world without one has no external signal that its optimization is draining a shared planetary resource. The monitoring body is not bureaucracy. It is the signal the optimizer needs to avoid catastrophe.

Governing the commons before AGI arrives is not idealism. It is the minimum precondition for AGI that does not destroy them. The Fair Framework is that precondition. The window is open now.

VII What Must Happen and When: The Sequencing Imperative

This paper does not propose slowing AGI development. It argues that AGI deployed into ungoverned planetary commons is structurally guaranteed to accelerate the exact governance failures this series has documented — and that the solution is to govern the commons before the optimizer arrives.

TIMEFRAME	REQUIRED ACTION	FAIR FRAMEWORK INSTRUMENT
2026–2027	Political mobilization; scientific consensus publication; this series submitted to UNEP, UNESCO, and WMO for formal review	The Fair Framework Series as policy submission vehicle
2027–2028	Monitoring body establishment; baseline data collection mandated; international scientific working groups convened per system	PEBO, GOAMN, GMBMN, GAMRS launched under existing UN architecture
2028–2030	Treaty negotiation initiated for binding governance of highest-risk commons: orbital space, atmospheric geoengineering, aquifer systems	ISTMA, IGGT, TAGP negotiation tracks opened
2030–2033	Binding treaty ratification; cumulative impact assessment protocols mandatory for all AGI-assisted large-scale interventions in planetary commons	Full Fair Framework institutional architecture operational prior to median AGI timeline

> *"You cannot manage what you cannot measure. You cannot measure what no institution is mandated to see. And you cannot align an optimizer with a commons that does not yet know how to see itself."*

VIII Conclusion: The Window

The seven papers that precede this one documented the same structural failure across seven planetary systems: humanity is running planet-scale uncontrolled experiments on its own life-support infrastructure, in isolation, without aggregate oversight, and without the institutional architecture to detect irreversible thresholds before they are crossed.

This paper adds a single, urgent dimension to that diagnosis: the optimizer is coming. Within the credible expert range of AGI timelines, a system capable of acting on every one of these systems simultaneously — at superhuman speed, with no inherent regard for commons that no institution has been mandated to protect — will be deployed by actors whose individual rationality will be, as it has always been, individually rational and collectively catastrophic.

The Fair Framework is the precondition. Not for AGI alignment — that is the necessary work of the AI safety community. But for the aligned AGI to have a world of governed commons to operate within. Alignment without commons governance is an aligned optimizer in an ungoverned commons. The math resolves the same way it always has. At AGI speed.

The window is open now. The physics will not wait. The optimizer will not wait. The commons — seven of them, documented across this series, each accumulating irreversible damage — will not wait.

This is how we help save our world: by building the measurement architecture that makes collective action possible — before the most powerful optimizer in human history arrives in a world that has none.

Govern the commons first. The optimizer is already on its way.

REF

References

1. Metaculus. (February 2026). When will the first general AI system be devised, tested, and publicly announced? Metaculus Forecasting Platform.
2. Todd, B. (February 2026). Shrinking AGI Timelines: A Review of Expert Forecasts. 80,000 Hours.
3. AI Futures Research. (April 2026). Q1 2026 Timelines Update. aifutures.org.
4. Future of Life Institute. (December 2025). AI Safety Index Winter 2025.
5. Creutzig, F. et al. (2026). Governing artificial intelligence for planetary health. The Lancet Planetary Health.
6. Kokotajlo, D. et al. (April 2025). AI 2027. AI Futures Research.
7. Kulveit, J. et al. (2025). Gradual Disempowerment. Alignment Forum / Oxford.
8. Sahoo, S. (2025). Policy Myopia as a Mechanism of Gradual Disempowerment in Post-AGI Governance. Cambridge AI Safety Hub.
9. Gros, C. (2023). Generic Catastrophic Poverty When Selfish Investors Exploit a Degradable Common Resource. Royal Society Open Science.
10. IPCC. (2022). Sixth Assessment Report: Impacts, Adaptation and Vulnerability. Cambridge University Press.
11. Larsen, O.F.A. & van de Burgwal, L.H.M. (2021). On the Verge of a Catastrophic Collapse? Frontiers in Microbiology.
12. Belfer Center for Science and International Affairs. (December 2025). Governing Outer Space: A Conference of the Parties for the Outer Space Treaty.
13. Global Commission on the Economics of Water. (2023). Turning the Tide: A Call to Collective Action.
14. UNEP. (2023). One Atmosphere: An Independent Expert Review on Solar Radiation Modification Research and Deployment.

15. Fair, R.T. Jr. (2026). The Fair Framework: Papers I–VII. Independent Global Systems Policy Research.

Robert Thomas Fair Jr. Nephew of Bill Fair — Co-Founder of FICO

The Fair Framework: Eight Papers in Planetary Governance and the AGI Prerequisite | May 2026

FICO did not eliminate financial risk. It made risk visible, measurable, and therefore manageable at scale. The Fair Framework applies the same instinct to planetary systems — before the optimizer arrives.

© 2026 Robert Thomas Fair Jr. All Rights Reserved. Independent Global Systems Policy Research