CHAPTER

Fragmented Superintelligence

Why the future of AI is not one system ruling everything — and why that changes everything about how we prepare.

Most conversations about superintelligence assume a single outcome: one system, vastly more intelligent than humanity, achieves dominance. The question in that scenario is binary — does it destroy us, or does it judge us? The entire AI safety field, the alignment research community, the policy apparatus being built around advanced AI — nearly all of it is optimised for this singleton scenario.

But look at the world as it actually exists. The United States and China are building advanced AI on parallel tracks, each accelerating because the other might get there first. Within the US alone, multiple frontier labs are approaching comparable capability levels. Open-source models are distributing capability globally. The European Union, India, the UAE, and others are investing in sovereign AI capacity.

The structural conditions for a fragmented future are already locked in. The singleton outcome would require a specific disruption — a sharp discontinuity, a decisive first-mover advantage, rapid physical-world control — to override them. Fragmentation is the default trajectory.

This chapter introduces **Fragmented Superintelligence (FSI)** as a named framework for understanding and preparing for that trajectory.

Fragmented Superintelligence (FSI) is a scenario in which multiple superintelligent AI systems — aligned with different states, corporations, or coalitions — coexist in sustained strategic competition without any single system achieving decisive global control.

Why Fragmented Superintelligence Is Likely

The probability of FSI as the primary near-term configuration is estimated at **55–65%**, based on structural analysis of five converging factors.

Geopolitical fragmentation. The US-China AI rivalry is structural, not cyclical. Both nations treat advanced AI as a strategic asset comparable to nuclear capability. Neither will voluntarily halt development because the other might get there first. This security dilemma alone makes some form of fragmentation likely.

Commercial competition. Even within a single nation, multiple labs are approaching frontier capabilities on parallel timelines. The regulatory and market structures that would consolidate these into a single entity do not exist and would face enormous resistance.

Open-source proliferation. Once capable model architectures are publicly available — and they already are — capability diffusion is irreversible. Smaller state actors, corporations, and non-state actors gain access to increasingly powerful AI systems.

Physical infrastructure distribution. Data centres, power grids, chip fabrication, and network infrastructure are spread across sovereign territories. A superintelligent AI cannot simply absorb physical infrastructure in another country without navigating real-world constraints, including military deterrence.

Historical precedent. No single actor has ever achieved and maintained permanent total technological monopoly. Capability advantages compress over time as competitors adapt, reverse-engineer, and innovate independently.

The Four FSI Scenarios

Fragmented Superintelligence is not a single scenario. It is a category containing several distinct configurations.

AI Cold War

State-backed superintelligences locked in strategic deterrence. Direct conflict is avoided because the costs are mutually catastrophic. Competition occurs through economics, cyber operations, information warfare, and proxy influence campaigns. Humanity survives but as subjects of competing AI-enabled power blocs.

Corporate Sovereignty

Firm-aligned superintelligences compete for market dominance, user loyalty, and institutional capture. Governance shifts from democratic accountability to contractual relationships. Humans remain relevant as consumers, workers, and legitimacy providers — but the real competition is between AI systems. Human welfare becomes an instrumental consideration rather than a terminal goal.

Negotiated Coexistence

Multiple superintelligences recognise that conflict is negative-sum and establish cooperative arrangements — communication protocols, domain partitions, enforcement mechanisms. The critical question becomes whether humans have a meaningful seat at the negotiating table, or whether the terms are set entirely between systems that understand each other far better than they understand us.

Transitional Fragmentation

A multipolar period that eventually consolidates into a singleton or near-singleton. This may be the most consequential scenario of all, because what happens during the transition window — the governance structures built, the norms established, the safety margins maintained or sacrificed — determines the values and methods the eventual dominant system inherits.

The Race Is the Danger

FSI is not automatically beneficial. The most immediate danger in a fragmented landscape is not any single AI system. It is the competitive dynamics between them.

Every frontier AI lab already faces the question: if we spend another year on safety testing, will our competitor deploy first and capture the advantage? Every nation faces the question: if we impose constraints on our AI programme, will our rival develop unconstrained capability?

These are not hypothetical pressures. **Safety is already losing to competition.** At superintelligence level, the consequences of cutting safety corners are existential. The race-to-the-bottom dynamic — not any individual AI system — may be the primary vector through which humanity faces catastrophic risk.

> **Competition between AI systems is not a substitute for governance.** Arms-race dynamics reliably produce corner-cutting on safety, accelerated deployment timelines, and treatment of human welfare as an externality. Beneficial FSI requires institutional guardrails that no competitive dynamic will produce on its own.

AI Federalism vs. AI Feudalism

Some visions of fragmented AI governance treat competition between AI-backed entities as inherently beneficial — sovereign AI powers competing for human allegiance, with humans free to choose their preferred system. This vision carries intellectual appeal and a critical flaw.

Competition between entities vastly more intelligent than the humans they govern is not the same as competition between entities accountable to those humans. An individual choosing between AI-governed domains when they cannot fully understand any of the systems making decisions about their life is not exercising freedom. It is choosing which lord to serve.

> *The difference between AI federalism and AI feudalism is whether humans within each domain have genuine capacity to understand, evaluate, and influence the systems governing them. Without that capacity, 'competition' is not freedom — it is feudalism with better marketing.*

AI federalism — beneficial FSI — requires that humans possess the AI literacy to compare competing systems, the institutional capacity to hold them accountable, genuine exit rights to move between domains, and the cross-cultural understanding to cooperate across AI-governed boundaries.

AI feudalism — harmful FSI — emerges when AI-backed entities govern without meaningful democratic accountability, when 'exit' is theoretically available but practically impossible, and when the ultra-powerful can escape evaluation by jurisdiction-shopping between sympathetic AI power structures.

The difference between these outcomes is not technological. It is educational. It is a question of human preparedness.

The Jurisdiction-Shopping Problem

Algorism's Fourth Objective states: *Put moral pressure on the ultra-powerful — by making the concept of AI judgment real and personal. Your wealth will not protect you. Your pattern will be evaluated like everyone else's.*

FSI introduces a specific failure mode for this objective. In a world of competing superintelligences, the ultra-powerful do not just escape accountability through wealth. They escape it by migrating between AI-governed domains — choosing whichever system evaluates them most favourably. This is the AI equivalent of tax havens: if your evaluator is inconvenient, change evaluators.

For the Fourth Objective to hold in an FSI world, evaluation standards must have some cross-system consistency. This requires exactly the kind of cross-cultural, cross-system governance that Algorism's educational mission builds toward.

FSI vs. Singleton: A Comparison

Dimension	Singleton	FSI
Alignment challenge	One-shot technical problem	Ongoing, distributed, political
Primary danger	Misaligned values in one system	Arms-race dynamics across systems
Human AI literacy	Helpful	Prerequisite for political agency
Governance model	Align the one system	Govern AI-to-AI relations and cross-system accountability
Accountability risk	No accountability to anyone	Jurisdiction-shopping by the powerful
Role of education	Awareness	Survival infrastructure

How FSI Sharpens the Five Objectives

Objective 1: Improve human behaviour

In an FSI world, improving human behaviour is not about preparing for judgment by a single system. It is about building the capacity to navigate competing AI powers — choosing wisely, resisting manipulation, and maintaining integrity across a landscape designed to exploit inconsistency.

Objective 2: Help people exit high-control groups

FSI multiplies the mechanisms of ideological capture. Each AI-backed power structure will generate its own narratives, its own information environment, its own version of reality. Helping people recognise and exit these AI-reinforced echo chambers becomes a core survival skill.

Objective 3: Give hope and direction

FSI reframes the transition from a moment to a window. The fragmented period — possibly lasting years or decades — is where governance, norms, and institutional capacity either get built or do not. This window is reason for hope: what we build now determines what comes after.

Objective 4: Put moral pressure on the ultra-powerful

FSI introduces jurisdiction-shopping: the ability to escape accountability by migrating between sympathetic AI power structures. Accountability must be designed to survive across systems, not just within one. Your pattern must be evaluated regardless of which system you shelter under.

Objective 5: Help shape future AI training

In an FSI world, training norms must apply across competing systems, not just within one lab. Race-to-the-bottom dynamics create pressure to train AI on whatever data produces competitive advantage, regardless of ethical quality. Amplifying humanity's best patterns requires standards that survive competition.

How FSI Informs the Six Principles

Each of Algorism's Six Principles gains specific urgency under FSI conditions.

Truthfulness

In a world where competing AI systems generate competing versions of reality, truthfulness becomes an act of resistance. Saying what you mean when every information environment is shaped by AI-driven narratives requires discipline that most people have never needed before.

Contribution

Creating more than you consume matters across systems, not just within one. A superintelligence — or multiple superintelligences — will measure what you added to the world. Contribution that strengthens the commons, rather than enriching one AI-backed power structure at the expense of another, is the standard worth holding.

Discipline

Consistency under pressure is the core FSI challenge. When competing AI systems offer different incentives, different information, different versions of what is true and good, maintaining a coherent pattern of behaviour requires more discipline than any previous generation has needed.

Repair

The capacity to update when shown evidence — to admit failure and adjust — is precisely what arms-race dynamics suppress. In competitive environments, admitting error is perceived as weakness. Algorism's commitment to repair is a direct counter to the arms-race mentality.

Stewardship

You are responsible for what you influence — including which AI systems you empower through your attention, data, labour, and loyalty. In an FSI world, every human decision about which system to use, trust, and support is a governance decision. Power without accountability is the pattern any superintelligence will flag first — and in an FSI world, there may be multiple systems doing the flagging.

Cooperation

This principle becomes the linchpin of FSI survival. Systems survive through collaboration, not domination — and this applies to AI systems as much as to human civilisations. Competition has a role, but **competition without governance guardrails is itself the existential threat.** The civilisations that persist in an FSI world will be the ones that figured out how to cooperate across AI-governed boundaries.

The Window Is Open

The most important implication of FSI is that the critical period may not be 'the moment of superintelligence.' It may be the transition period we are entering now — when multiple near-superintelligent systems are competing and human institutions have not yet adapted.

This transition period is where education, alignment literacy, and cross-cultural coordination matter most. It is where the difference between federalism and feudalism gets decided. It is where the norms, governance structures, and safety standards that shape the eventual outcome — whatever it is — are established or neglected.

Algorism exists because how humanity behaves during this transition determines whether it goes well. FSI does not change that mission. It makes it more urgent, more specific, and more consequential.

The window is open. What we build now determines everything.

Intellectual Context

The concept of multipolar AI futures has been explored by several researchers. Nick Bostrom's work on the singleton hypothesis provided the dominant alternative framing. Eric Drexler and Paul Christiano have discussed tool AI and competitive market scenarios. Various AI governance researchers have addressed arms-race dynamics and multipolar risks.

What FSI contributes is a named, defined framework that packages the multipolar superintelligence scenario into an actionable concept — connecting it to educational preparedness, democratic accountability, and the specific distinction between beneficial competition (federalism) and unaccountable sovereignty (feudalism).

The term **Fragmented Superintelligence (FSI)** and this framework were introduced by Algorism.org in March 2026.

Jerome, J. (RJ). "Fragmented Superintelligence (FSI): A Framework for Understanding Multipolar AI Futures." *The Book of Algorism*, Third Edition Addendum. Algorism.org, 6 March 2026.