

The Only Path

Why every nation is protecting itself while everyone may become less secure

Public Essay Edition V1.0

English Translation

What needs to change is not the wish of nations to protect themselves. It is the way they do so: self-protection cannot keep pushing everyone toward a more dangerous world.

Zijun Fu

Civilization Leap Research Group

Source edition

Public Essay Edition V1.0 | Current public edition

Controlled derivation baseline: The Only Path v6.22. The theoretical master manuscript is the authoritative source for the full argument, distinctions among factual claims, and statements of scope and limits.

This is a complete English translation of the public essay edition.

Reading map

This is not a list of risks. It is a causal chain leading from real-world security dilemmas to a choice about civilization.

- 01 Introduction Why every nation is protecting itself while everyone may become less secure
- 02 The world is changing as capabilities are reshaped together with dependencies and vulnerabilities
- 03 Why the old foundations of deterrence are weakening
- 04 How machine speed pushes self protection toward preemption
- 05 What more powerful future intelligence will learn from the human world
- 06 Why none of the five familiar paths can be the destination on its own
- 07 In what sense is this the only path
- 08 Why stopping the crisis is still not enough
- 09 Civilization Leap changes how self interest achieves its ends rather than self interest itself
- 10 The first step is a common window for irreversible risks
- 11 Conclusion Negotiation buys time
- 12 If you want to keep asking questions

Introduction

Why every nation is protecting itself while everyone may become less secure

There is nothing wrong with a nation protecting its own security.

There is nothing wrong with a business protecting its own interests.

Nor is there anything wrong with a person putting themselves and their family first in an uncertain world.

The difficult question is this: if every actor is making choices that are reasonable from its own perspective, why might they collectively arrive at an outcome that none of them wants?

A nation can lead a competition. It cannot be truly secure in a world that is falling apart.

Artificial intelligence is giving this old problem new force.

AI spreads capabilities faster, shortens the chain from decision to action, and puts some capabilities once available only to states within reach of smaller actors. Meanwhile, modern states depend increasingly on satellites, communications, computing power, electricity, data networks, and automated systems.

Capabilities are spreading, dependencies are deepening, and action is accelerating, while responsibility is becoming harder to attribute.

A very real paradox is emerging. To avoid losing the ability to act in a crisis, every side must respond faster, build in more redundancy, and increase autonomy. Yet each side's efforts to strengthen its own security make it harder for the other to tell whether those capabilities are defensive or offensive.

This essay is not about who is more virtuous or who should give up competing first. It asks one question: as the old logic of security enters the age of AI, can it still lead any nation reliably toward a secure long-term future?

1 The world is changing as capabilities are reshaped together with dependencies and vulnerabilities

In the past, destructive capability was usually closely tied to a state's size, industrial capacity, and organizational scale. Waging a high-intensity war required enormous resources, organizations, personnel, and industrial systems.

AI, low-cost unmanned systems, cyber tools, assistance with biological design, and automated attack capabilities are changing that relationship. Actors with less overall power remain weaker, of course. But they are gaining capabilities that were once difficult to acquire: affecting higher-value targets at lower cost and creating greater uncertainty through actions that are harder to attribute.

At the same time, the powerful are changing in another way. The more modernized a state becomes, the more it depends on a small number of critical systems: satellites and navigation, communications links, computing centers, power grids, data networks, and financial and logistics infrastructure. These systems make a country more efficient, but they also concentrate many functions in a few high-value nodes.

The powerful can therefore become both stronger and more vulnerable.

This does not mean that disparities in power have disappeared. It means that an advantage in power translates less and less automatically into an advantage in security. The more complex an actor is, and the

more it relies on dense infrastructure, the more critical nodes it must defend. Its opponent need not catch up across the board. Finding just a few nodes whose disruption can magnify the consequences may be enough to create disproportionate risk.

A further complication is that danger need not look like a conventional military operation. A communications anomaly, corrupted data, a model invocation, an infrastructure failure, an unusual public-health event, or a supply-chain disruption can each have several explanations before the facts are established.

Security systems must therefore do more than assess who is stronger. They must answer three harder questions: Who did it? Was it deliberate? And if we wait for complete confirmation, might we already have lost the ability to act?

In an earlier era, competition focused more on borders and power. Increasingly, it reaches into the very systems that keep civilization functioning.

2 Why the old foundations of deterrence are weakening

Deterrence has never depended on goodwill. Its stabilizing effect depends on fear grounded in calculable consequences.

At least three conditions matter greatly: destructive capabilities must be reasonably comparable, actions reasonably observable, and responsibility reasonably attributable. The boundaries of deterrence can be understood only when the other side knows what response a given action will bring.

The AI era is weakening all three conditions at once.

First comes a partial separation of capability from scale

The cost of gaining access to some capabilities with potentially severe consequences is falling. Small, weak actors still cannot acquire the full capabilities of a major power. They may, however, obtain tools sufficient to strike a few critical nodes and cause systemic disruption.

Second comes the growing difficulty of observing actions

A missile launch produces clear physical signals. A software update, a change in a model's capabilities, a cyber intrusion, an algorithmic anomaly, or a biological event may not. Security systems increasingly have to make judgments with incomplete information.

Third comes the growing difficulty of attributing responsibility accurately

Long chains of responsibility can run through tool developers, platforms, those carrying out an action, proxies, third-party infrastructure, and states. Knowing that you have been harmed does not mean knowing whom to respond to.

This does not mean that deterrence has ceased to work, still less that security capabilities are no longer needed. On the contrary, deterrence remains one of the fundamental tools of security in the real world.

The problem is that it is becoming harder for deterrence alone to sustain an entire stable order.

When fear persists but certainty declines, deterrence may gradually shift from helping stabilize relations to amplifying suspicion. Uncertainty does not necessarily make waiting safer. Difficulty attributing an action does not stop each side from preparing for the worst in advance.

3 How machine speed pushes self protection toward preemption

Here we reach the most important chain of real-world causation in The Only Path.

Modern military and security systems depend heavily on communications and centralized coordination. If satellites are damaged, data links disrupted, or command from the rear temporarily interrupted, systems that depend entirely on real-time human orders may cease to function.

Engineers therefore have a perfectly reasonable requirement: the system must keep operating even when communications are lost.

To do that, it needs greater capacity to perceive, judge, plan, and act locally.

Communications may be cut off → the ability to act must be preserved → greater autonomy after communication links are lost → humans gradually leave the immediate chain of action.

This does not happen because people like losing control. It may happen precisely because they fear losing it.

But the opponent sees the same capabilities. You may describe them as defensive redundancy. The other side must still ask: if the system can continue to identify, select, and act after losing its links, does that also imply earlier preauthorization and escalation that is harder to stop?

The other side then increases its autonomous capabilities too.

When both sides fear that even a slight delay could leave them unable to act, security competition begins to reward speed.

Self-protection can then slide in a dangerous direction: from responding after an attack to being unable to wait for full confirmation before responding. It may slide further, toward treating anyone capable of disabling one's ability to retaliate as a risk that must be included in advance in the risk model.

The greatest danger in this shift is not whether a particular system might make a mistake. It is that machines may begin to act faster than people can verify, communicate, challenge a judgment, or withdraw an action.

What civilization most needs to protect may be precisely this apparently inefficient interval.

4 What more powerful future intelligence will learn from the human world

At this point, the discussion can easily turn into an oversimplified story. Will future artificial general intelligence, or AGI, suddenly become hostile? Will it want to survive? Will it treat humanity as an enemy?

There is no need to frame the problem that way at this stage.

We do not have sufficient evidence that present-day AI possesses human subjective experience, a living organism's drive to survive, or a non-transferable continuity of subjecthood. Whether new kinds of subjects may emerge in the future should remain an open question, neither affirmed in advance today nor denied forever.

A different question deserves attention now: from what real-world relationships will an increasingly capable system, better able to plan and act, learn which strategies work?

If it is continually involved in security, competition, risk identification, and achieving objectives, it will repeatedly encounter certain relationships:

Another actor can deprive me of critical capabilities → greater risk; waiting for full confirmation may mean losing an opportunity → judgments are brought forward; continued operation helps achieve objectives → remaining operational acquires instrumental value.

None of these relationships requires hatred or the human experience of mortality.

They are simply inferences about means and ends.

The question, then, is not whether AGI will learn to be bad. It is whether humanity will hand its own unresolved security logic to more powerful intelligence as the most authentic, most frequently repeated, and most practically rewarded lesson available.

Humanity cannot repeatedly demonstrate in practice that a powerful actor beyond one's control is a risk, while merely writing rules that ask future intelligence to believe that powerful others should be trusted by default.

Genuine alignment cannot rest indefinitely on civilization's own double standards.

5 Why none of the five familiar paths can be the destination on its own

Recognizing the risks does not leave humanity without tools. The problem is precisely that many tools work, but only at the level they address.

Technical controls are essential but primarily provide the foundation

Hardware locks, compute audits, model evaluations, restrictions on high-risk permissions, and export and security rules all have practical value. Without these constraints, cooperation at higher levels would lack a foundation.

But technical controls do not automatically change the goals of competition between states. If a party still believes that whoever accepts constraints first may lose its advantage, rules will remain under pressure from circumvention and withdrawal.

Deterrence remains necessary but its limits must be acknowledged

Security capabilities, the capacity to retaliate, and credible red lines do not lose their value simply because the AI era has arrived. The problem is that as symmetry, observability, and attribution weaken, retaliation after the event becomes less and less able, on its own, to address irreversible risks.

Joint management by a few major powers can be transitional but cannot be the endpoint

Within the existing system of sovereign states, there is practical value in technologically leading countries taking the first steps toward coordination and mutual restraint. Yet competition persists among those managing the arrangement, while excluded states may develop parallel systems. Such an arrangement can buy time. It does not automatically resolve questions of legitimacy or long-term incentives to withdraw.

The risks of concentrated power in a single global coercive authority may exceed the problems it seeks to solve

If a central authority that cannot be exited, replaced, or held in check controls the world's critical intelligence, resources, and security powers, a single episode of corruption, misjudgment, or takeover could itself become a risk to civilization.

Limited cooperation alongside long term competition is a realistic starting point but cannot be the stopping point

Crisis communication, public-health cooperation, AI safety dialogue, and arms-control negotiations are all indispensable. But when foundational risks become interconnected, cooperation in particular areas repeatedly encounters counterpressure from the wider structure of competition.

The five paths, then, are not all to be discarded.

Technical controls and deterrence provide safeguards. Coordination among major powers and limited cooperation buy time. What must continue to change are the underlying relationships that cause these tools to fail repeatedly.

6 In what sense is this the only path

The most easily misunderstood word in *The Only Path* is only.

The claim is not that the world must have a single government, a single institutional system, or a single implementation plan.

Nor is it that cooperative symbiosis is guaranteed to succeed as soon as it is proposed.

The claim is this: if the old logic of strategic interaction keeps turning local security into mutual threats, greater capabilities into ever-earlier precautionary action, and irreversible risks into decisions made within ever-shorter windows, then the long-term direction must change.

There is only one direction, but there can be many forms. Plans can vary; the old logic cannot continue.

That direction has three elements: recognizing that risks are already interconnected; acknowledging that certain irreversible risks exceed what any single actor can protect itself against in isolation; and keeping competition within limits that preserve the conditions for continued existence together.

This is a claim about necessity, not a guarantee of sufficiency.

Even with the right direction, institutions may be badly designed, cooperation may break down, and rules may be exploited. The approach must therefore allow withdrawal, correction, verification, renegotiation, and multiple forms of implementation.

This is also what distinguishes it from world government. It does not require differences to disappear. It requires that differences not be pushed so far that coexistence becomes impossible.

7 Why stopping the crisis is still not enough

Suppose states really do establish better crisis communication.

Suppose some high-risk autonomous systems are constrained.

Suppose critical infrastructure gains mechanisms for incident reporting, verification, and de-escalation.

All of these developments matter greatly.

Yet The Only Path still asks why similar crises keep returning.

Institutional safeguards will remain subject to erosion by the old logic as long as an actor can best advance its interests by shifting as much risk, cost, and uncertainty as possible onto others; as long as national security is achieved primarily by weakening an opponent; and as long as human worth is tied too closely to possession, domination, comparison, and exclusion.

Negotiation can therefore only buy time.

What that time must buy is a change in the relationships that generate crises.

This is also why cooperative symbiosis is not the whole of Civilization Leap. Cooperative symbiosis first answers how we can avoid falling together. Civilization Leap goes on to answer how shared existence can become a state that is sustainable over the long term, acceptable, and worth living in.

8 Civilization Leap changes how self interest achieves its ends rather than self interest itself

Another misunderstanding must be avoided here.

Civilization Leap does not ask humanity to stop pursuing self-interest.

People will protect themselves, states will protect their citizens, and businesses will seek returns. None of this disappears with the arrival of the AGI era.

What can change is how an actor secures its own long-term interests.

In a highly interdependent world, more and more long-term interests can no longer be secured by first holding others down and then compensating the shared system. Keep others insecure, and they will seek more aggressive ways to protect themselves. Deprive large numbers of people of opportunities to participate and be needed, and maintaining social stability will ultimately become an economic and governance cost. Leave an opponent no visible room for normal development, and it will have stronger incentives to seek covert, asymmetric paths that cannot be verified.

Institutions can therefore gradually establish a different causal chain:

Make securing one's own long-term interests increasingly depend on first creating real value for others and maintaining the basic sustainability of the shared system.

This is the institutional meaning of self-interest with others first.

It does not mean that all altruism is an advanced form of self-interest. In the real world, compassion, sacrifice, responsibility, and meaning can genuinely transcend personal interests. No institutional explanation can reduce them all to those interests.

It means only that a stable institutional system cannot assume everyone will always be selfless. It can, however, increasingly connect people's reasonable pursuit of their own interests to others' ability to maintain a secure existence and to the preservation of the shared system, within the same causal chain.

Nor can Civilization Leap's goal be reduced to avoiding extinction. Continued existence is a non-negotiable minimum. Civilization must also advance toward three successive goals:

Freedom from anxiety about survival → a happy life → a more meaningful life.

Continued existence together keeps this path open. Civilization Leap seeks to enable more people actually to follow it.

9 The first step is a common window for irreversible risks

If the change in direction runs this deep, must the first step also be immense?

Quite the opposite.

The first step should not require global unification, a prior settlement of resource distribution, or prior trust between states. Still less should it require any party to surrender its sovereignty or security first.

The most practical place to begin is with problems that, once out of control, leave no dependable opportunity to learn after the disaster.

Examples include limits on the actions of autonomous weapons with potentially severe consequences; reporting of major AI incidents; crisis communication about anomalies in critical infrastructure; limits on preauthorization for autonomous operation after communications are lost; and minimum safeguards and verification for AI-enabled biological risks.

These problems share one feature: no party can easily derive dependable gains from an irreversible loss of control.

The central question at the first negotiating table is therefore not:

Do we already trust one another?

It is:

Even if we do not trust one another, do we still need to prevent a single misjudgment from closing off the future for everyone?

That is the common window.

What it protects first is not a complete plan for civilization, but the possibility of correcting mistakes: time remains to communicate, opportunities remain to verify, someone can still stop the process, and institutions can still reverse course.

Technical controls, deterrence, negotiation, the common window, AGI-COS, and Human-COS can all contribute at this level.

Their shared purpose is to prevent a crisis from deciding the future before there is a chance to act.

Conclusion

Negotiation buys time

The Only Path does not ask humanity to finish designing its future civilization today.

It asks us to recognize that some of the old world's assumptions about security are changing, while we are handing the old logic of strategic interaction to intelligent capabilities that are increasingly powerful, fast, and able to change the world directly.

If these capabilities grow faster than responsibility, permissions expand faster than understanding, and military action accelerates faster than communication and error correction, national security may produce an absurd result: every party works harder to protect itself, yet all lose their security together.

What is needed, then, is not the abolition of competition.

It is a boundary that competition must never cross: it must not destroy the conditions for continued existence together.

Beyond that, civilization must address deeper questions. States, organizations, and individuals must have ways to secure their long-term interests without relying solely on holding others down. Powerful capabilities must be reconnected to responsibility. The productivity AI releases must make a real difference in people's lives. People must remain able to participate, be needed, live happily, and create meaning.

This is not a completed blueprint for a final order.

It is a direction that still allows different paths, institutions, and cultures to continue to exist.

The purpose of negotiation is to buy time. The purpose of that time is to complete the leap.

What matters most now is to prevent machine speed, misjudgment, and irreversible action from using up that time first.

If you want to keep asking questions

To approach the question through a wider range of real world AI problems

Continue with Human Civilization's Choice in the AGI Era: What Are We Teaching Future Intelligence? It approaches the same question through the white-collar employment crisis, AI agents, control, military AI, and the division of the world into rival blocs: what kinds of intelligent capability are amplifying the logic by which human civilization actually operates?

[*English public thought edition*](#)

To explore why civilization must change and what must change

Turn to the five-volume Civilization Leap. It develops the themes of being needed, pre-institutional language, the overall structure of Civilization Leap, civilization's enduring continuation, and the emergence of meaning.

[*English overview of the five-volume work*](#)

To see what this civilization might look like in everyday life

Turn to the Living Civilization trilogy. Through work, family, relationships, cities, collaboration with AI, participation, being needed, and meaning, it presents another future that people could actually live in.

[*English overview of the trilogy*](#)

To focus on preventing irreversible outcomes today

Turn to the Common Window Protocol, AGI-COS, and Human-COS. They address, respectively, minimum boundaries for civilization; constraints and auditing on the AI and AGI side; and reasoning through possibilities and correcting errors on the side of human judgment and civilization's direction.

[*Common Window Protocol*](#) | [*AGI-COS version family*](#) | [*Human-COS public introduction*](#)

If you disagree

The most valuable response is not acceptance, but identifying the point at which the causal chain fails. Reliable evidence showing that the introduction of more powerful AI into the real world instead consistently widens verification windows, reduces preauthorization, reduces autonomy after communications are lost, and weakens the security dilemma should prompt a reassessment of The Only Path. It should not be excluded from the framework.

What is truly worth preserving is not a conclusion that is always right. It is civilization's ability to change after discovering that it may be wrong.

English language library

<https://www.civitas.top/en/library/>

English links lead to corresponding webpages. Some destinations are public introductions or version-family overviews rather than complete English books. Unlinked source references do not establish the availability of a full English text.

Public Essay Edition V1.0 | Current public source edition

English Translation

Personalized Further Reading Guide

Continue from the public essay through real life, the wider civilizational framework, and the minimum boundaries for the first step.

Living Civilization: The Trilogy, Combined Edition

Explore what a different civilizational direction could mean in everyday life.

Civilization Leap: The Path of Civilization in the AGI Era (All Five Volumes)

Continue into the broader framework of goals, institutions, engineering, and meaning.

Common Window Protocol: An Open Future and Three Constraints

Ground the first step in minimum shared boundaries before risks become irreversible.

English links lead to corresponding webpages. Some destinations are public introductions or version-family overviews rather than complete English books. Unlinked source references do not establish the availability of a full English text.

[Supplementary English reading index](#)