AGI-COS 2.1 Ministerial Brief（部长速读版）。

这是现实世界中最关键版本：

- 政府部长
- 国家安全顾问
- 国际组织负责人
- 科技监管高层

真正会阅读的，不是完整版，而是这种 8 页以内决策简报。

本版本严格按 西方政策系统阅读逻辑设计。

AGI-COS 2.1

Global AGI Stability Framework

Ministerial Brief — Executive Policy Version

Classification: Policy Concept Brief
Reading Time: ~10 minutes
Purpose: Decision Awareness & Pilot Consideration

---

1. Executive Overview

Artificial General Intelligence (AGI) introduces a new category of global risk:

Not technological failure —
but coordination failure under accelerating intelligence.

Current governance systems were designed for slower technological eras.

AGI-COS 2.1 proposes a lightweight international coordination infrastructure allowing states and industry to manage shared risks without sacrificing sovereignty or innovation.

---

2. Why This Matters Now

AGI changes three structural conditions simultaneously:

1. Speed

AI capability evolves faster than policy response cycles.

2. Scale

Single technical events can trigger cross-sector global effects.

3. Irreversibility

Some outcomes may not be correctable once deployed.

The strategic question is no longer:

Can AGI be developed safely?

but:

Can humanity coordinate fast enough?

---

3. The Core Policy Problem

Today's risks resemble a classic coordination dilemma:

- No country wants unsafe AGI.
- No country wants to slow alone.
- Competition incentives dominate cooperation.

Result:

Collective instability despite rational actors.

AGI-COS addresses coordination — not control.

---

4. What AGI-COS 2.1 Is (and Is Not)

AGI-COS IS:

- A shared analytical framework
- A coordination infrastructure
- A voluntary stability mechanism

AGI-COS IS NOT:

- A global regulator
- A treaty enforcement body
- A restriction on national capability

Human decision authority remains fully sovereign.

___

5. How the Framework Works

AGI-COS introduces five operational components.

___

A. Shared Risk Awareness

Common indicators enable aligned understanding:

- System Stress Index (SSI)
- Irreversibility Risk Index (IRI)
- Trust Calibration Score (TCS)
- Civilizational Time Index (CTI)

Purpose:
Create shared situational awareness without sharing sensitive technologies.

___

B. Independent Analytical Layer

AI systems provide:

- scenario modeling,
- risk forecasting,
- uncertainty disclosure.

AGI assists decisions but never makes them.

___

C. Reversible Intervention Model

Instead of large regulatory moves:

- small pilot actions,
- reversible adjustments,

- continuous learning.

Reduces political and economic risk.

___

D. Crisis Coordination Protocol

When risk thresholds rise:

- rapid analytical exchange,
- synchronized communication,
- voluntary alignment of responses.

No compulsory authority involved.

___

E. Global Learning Loop

Incidents and responses become shared lessons, preventing repeated systemic failures.

___

6. Why Countries May Participate

AGI-COS is designed to be strategically neutral.

Benefits include:

National Security

Reduced accidental escalation risk.

Economic Stability

Predictable innovation environment.

Technological Leadership

Participation shapes future global norms.

Sovereignty Protection

No external decision authority.

---

7. Adoption Strategy (Realistic Path)

AGI-COS does not require immediate global agreement.

Implementation follows a proven pathway:

Phase 1 — Pilot Cooperation (1–2 years)
- voluntary participants
- non-sensitive domains
- shared metrics testing

Phase 2 — Operational Coordination (3–4 years)
- incident communication standards
- analytical interoperability

Phase 3 — Institutional Integration (5+ years)
- integration into existing international frameworks.

---

8. First-Mover Protection

A simultaneous activation model ensures:

No participating state faces competitive disadvantage.

Commitments activate only after sufficient participation threshold is reached.

---

9. Strategic Value for Decision-Makers

AGI-COS offers governments a rare policy position:

Act early without committing to heavy regulation.

It enables preparation without escalation.

In strategic terms:

It buys decision time.

___

10. Immediate Next Steps (Low Commitment)

Decision-makers may consider:

1. Observing pilot discussions
2. Participating in technical workshops
3. Assigning liaison experts
4. Evaluating compatibility with national AI strategy

No binding commitment required.

___

Final Note

The defining risk of the AGI era is not intelligence itself, but whether humanity can coordinate under accelerating complexity.

AGI-COS 2.1 aims to ensure that:

technology advances,
human judgment remains central,
and civilization retains the freedom to choose its future.

AGI analyzes. Humans decide. Civilization continues.

___