

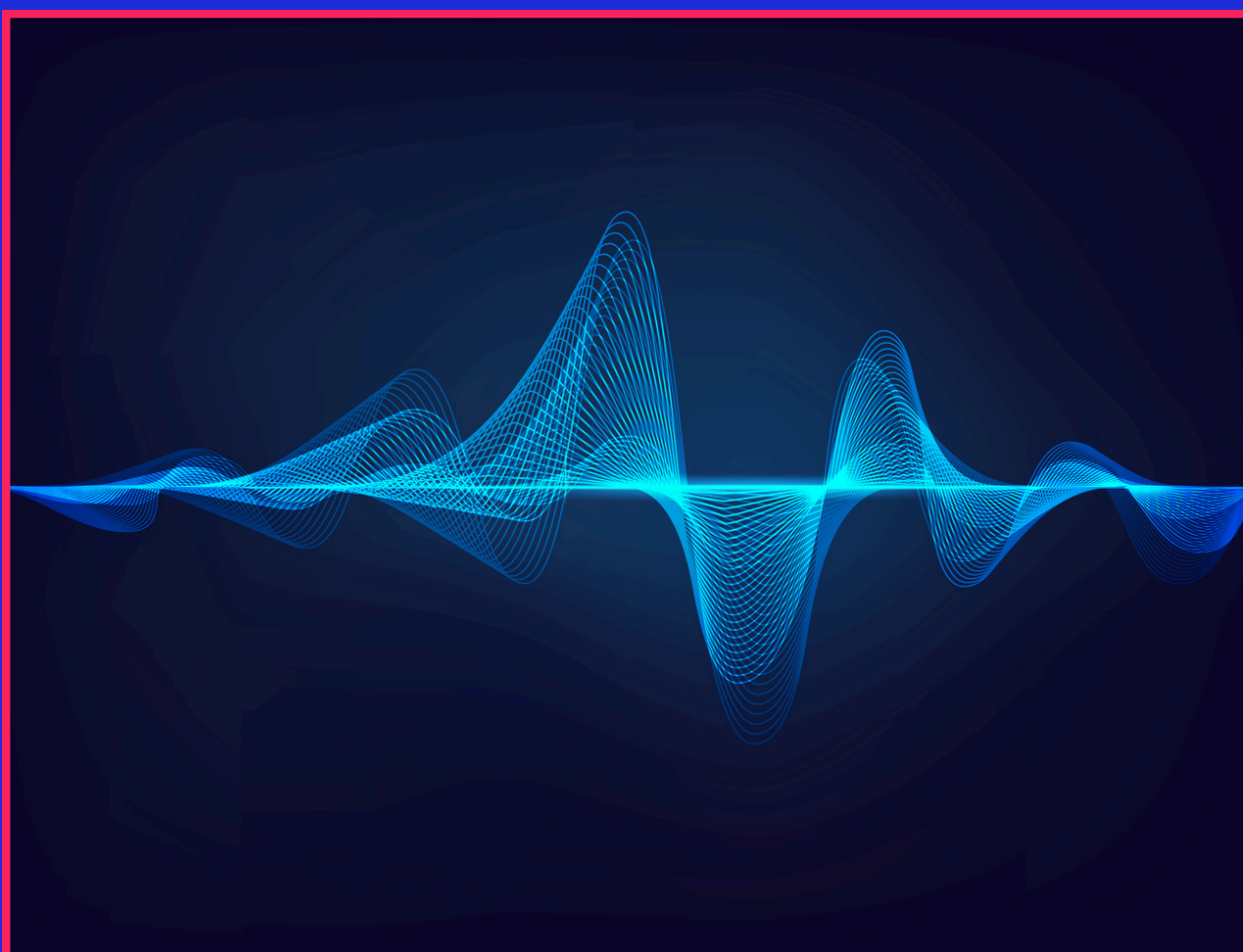
GAIGANow



# TONE MATTERS

2026

Richard Pagett



## **TONE MATTERS**

|                                                                                          |           |
|------------------------------------------------------------------------------------------|-----------|
| <b>1 Introduction.....</b>                                                               | <b>2</b>  |
| <b>2 Problems with Tone.....</b>                                                         | <b>2</b>  |
| 2.1 Disempowers Local Agency and Knowledge.....                                          | 2         |
| 2.2 Reproduces Neo-Colonial and Power Inequalities.....                                  | 3         |
| 2.3 Marginalises Southern Voices in Knowledge Production.....                            | 3         |
| 2.4 Leads to Misaligned or Ineffective Policies.....                                     | 3         |
| 2.5 Terminology and Framing Issues Perpetuate Stereotypes.....                           | 4         |
| <b>3 Some Case Studies.....</b>                                                          | <b>6</b>  |
| 3.1 Education: Universal Schooling Models & Curriculum Transfer, Sub-Saharan Africa..... | 6         |
| 3.2 Health: Structural Adjustment & Health System Reform.....                            | 7         |
| 3.3 Health: Global Health Research & “Helicopter Science”.....                           | 8         |
| 3.4 Climate Policy: REDD+ Forest Carbon Schemes.....                                     | 8         |
| 3.5 Climate & Development: “Good Governance” Conditionality.....                         | 9         |
| <b>4 Cross-Case Patterns.....</b>                                                        | <b>10</b> |
| 4.1 Frontier AI Risk Framed as a Universal Priority.....                                 | 10        |
| 4.2 Compute, Capability, and “Responsible Scaling” Frameworks.....                       | 11        |
| 4.3 AI Ethics & Safety Research as “Helicopter Science”.....                             | 12        |
| 4.4 “If You Don’t Regulate, You’re a Risk” Framing.....                                  | 12        |
| <b>5 Sector Cross-Cutting.....</b>                                                       | <b>13</b> |
| <b>6 Some Alternatives.....</b>                                                          | <b>14</b> |
| 6.1 Co-Production & Polycentric Governance.....                                          | 14        |
| 6.2 Capability-Based Risk Governance.....                                                | 14        |
| 6.3 Risk Pluralism.....                                                                  | 15        |
| 6.4 Infrastructure Justice & Access-First Governance.....                                | 16        |
| 6.5 South-Led Norm Entrepreneurship.....                                                 | 16        |
| 6.6 Procedural Justice in Global AI Institutions.....                                    | 17        |
| 6.7 Reflexive Safety.....                                                                | 17        |
| <b>7. Comparisons of Different Alternatives.....</b>                                     | <b>19</b> |
| <b>8. References.....</b>                                                                | <b>20</b> |

# 1 Introduction

What we have learned from climate change activism, during the last ten years, is that adopting a hectoring, a moralising, a patronising or a pejorative tone is unlikely to be persuasive. Consequently, when we wish to speak to others about advanced AI risk we need to be cognisant of our tone because not everyone perceives or prioritises AI risks in the same way.

There is a well-documented body of academic evidence and critique showing that when actors from the Global North draft papers, design policies, or set agendas using a *tone of prescription*, telling the Global South “what to do”, this can produce real problems across research, development practice, and policy.

The critique below is drawn from development studies, post-colonial and decolonial theory, and empirical studies of partnerships and knowledge production.

## 2 Problems with Tone

There are several problems identified in the literature.

These are, with evidence:

### 2.1 Disempowers Local Agency and Knowledge

Many academics argue that dictating solutions from the outside undermines the agency of people and institutions in the Global South.

- Research shows that Global North moral prescriptions often *mis-recognise the moral and political agency* of Global South actors and ignore local *geopolitical realities*, treating Southern actors as passive subjects instead of decision-makers (Wells, T. R., 2025); and
- Decolonial critiques highlight how typical research practices marginalise non-Western epistemologies<sup>1</sup>, sidelining local understanding, knowledge systems, and priorities (Gascoigne, J., *et al* (2025)).

#### **Why this matters?**

When policies or recommendations are externally imposed rather than co-produced, they risk being irrelevant, inappropriate, or even harmful because they do not reflect socio-cultural, political, or economic realities where they are applied.

---

<sup>1</sup> Epistemology is the branch of philosophy that studies the theory of knowledge (perception, testimony) and concepts such as belief, truth, and justification

## 2.2 Reproduces Neo-Colonial and Power Inequalities

A dominant theme in the literature is that the language of prescription reflects and reproduces structural power imbalances between the Global North and South.

- Historic and ongoing power asymmetries mean that Northern perspectives often frame Southern contexts as *problems to be solved*, perpetuating a legacy of *neo-colonialism* in research and development discourse (Rodríguez, A.L.G., 2022); and
- Even initiatives labelled “partnerships” frequently replicate unequal power dynamics, where Northern actors control funding, agenda setting, and decision-making (Downes, G., 2013).

### Why this matters?

Rather than challenging global inequalities, prescriptive tones can reinforce them by privileging Northern models as inherently superior or universal.

## 2.3 Marginalises Southern Voices in Knowledge Production

Global North dominance in academic publishing, research design, and agenda-setting means:

- Knowledge produced in the Global South is often “excluded, rejected, or side-lined”, despite its relevance and validity, (Kamruzzaman. P., 2024); and
- Debate on topics such as AI, development models, and digital ethics shows Southern perspectives are under-represented in global academic discourse, meaning the “solutions” promoted may miss critical considerations relevant to Southern contexts, e.g., AI ethics largely shaped by Northern frameworks (Roche, C. *et al* 2021).

### Why this matters?

What counts as “valid evidence” or “best practice” is shaped by dominant Northern paradigms, suppressing alternative Southern insights and reinforcing intellectual dependency.

## 2.4 Leads to Misaligned or Ineffective Policies

Prescriptive approaches often assume *one-size-fits-all* development models (e.g., modernisation theory, universal governance templates) without accounting for local conditions, leading to poor outcomes.

- ‘Development’ frameworks originating in the North may not align with priorities such as basic human needs, local governance capacities, or cultural frameworks in Southern contexts (Gerber, J-F., *et al* 2018); and

- In some fields (e.g., ecological research or education), critiques show that Northern research agendas dominate collaborations in ways that distort priorities and reproduce unequal knowledge hierarchies (Leal, J., *et al* 2022).

### **Why this matters?**

Policies that do not reflect lived realities or local priorities are less likely to succeed and can waste resources or cause unintended harm.

## **2.5 Terminology and Framing Issues Perpetuate Stereotypes**

Even apparently-neutral terms such as “Global South” can carry *implicit value judgments*. Critiques note that such terminology can over-simplify diverse contexts and inadvertently signal intellectual or developmental inferiority, which feeds back into prescriptive framing (Teixeira da Silva, J., 2021).

Pagett, R., (2025) noted that within the “Global South” there are distinct regions aligned with specific sets of development priorities in relation to AI benefit-sharing; namely, Sub-Saharan Africa, North Africa and the Middle East, South Asia, Latin America, the Caribbean and the Pacific States. It is also argued that China and India are not typical “Global South” countries and are better excluded from such a collective grouping.

As an alternative to “Global South”, an emerging descriptor is “Global Majority”, though even this can be problematic.

The term “Global Majority” was originally coined by Campbell-Stephens, R.M. (2021) in *Educational Leadership and the Global Majority - Decolonising Narratives*. This book introduced the term ‘Global Majority,’ as conceptualised within the context of school leadership. It examined the processes and impact over time of racially-minoritising up to eighty-five percent of the world’s population. The chapters illustrate how a decolonised cognitive reset from a minority to majority orientation moves practice from a place of subordination to one of agency and efficacy. By reconnecting the people of the Global Majority with their narratives and the social and historical linkages that they have always had, the book potentially contributes to a different global context; where interdependence is not driven by the economic greed of the minority, rather by the social and very human needs of the majority.

The challenge, though, is how to deal with China and India.

Global Majority is a collective term that refers to people who are Black, Asian, Brown, dual-heritage, indigenous to the global south, and/or have been racialised as 'ethnic minorities'. Globally, these groups represent approximately 80 - 85% of the world's population (with China and India), which shifts the narrative from a deficit (minority) to a position of strength and global belonging.

Yet through sheer geographical size, economic power, nuclear military might and advanced space technology, China and India are not in any sense a “minority”. They are both non-White which is the only real commonality they have with the rest of the Global Majority.

Of course, with a collective population of approximately 2.86 billion<sup>2</sup> (World Bank, 2024) these two countries do represent a significant portion of the world's total, roughly a third of the global population.

Yet, in practice, any reference to “Global Majority” should exclude both China and India. If China is excluded, yet India is retained, then the term becomes simply political, rather than analytical, and is conceptually vulnerable.

China often *is* excluded in practice, not usually explicitly. It tends to happen by silence or selective emphasis for several common, though weak reasons such as geopolitical discomfort. China is framed in Western discourse as a strategic rival and an authoritarian superpower and including China undermines the idea that the Global Majority are primarily disempowered. Another reason is moral framing, because some users of the term want “Global Majority” to imply: historical oppression, lack of global power and moral victimhood, in which China does not neatly fit. By including India, it can mask class, caste, and internal inequalities and feels conceptually strained when applied to citizens of such a powerful state; it is part of the G20. So, India should also be excluded from the collective term too.

So, “Global Majority” as a collective term in AI governance discourse does have agency, provided it excludes both China and India. These two countries should be at the table in their own right, not as some part of a collective, which they would undoubtedly dominate and influence.

### **Why this matters?**

Language shapes perceptions and policy: labels that reinforce binaries can entrench stereotypes rather than promote genuine, nuanced understanding.

### **In Summary**

Many of these critiques draw on longer traditions in critical theory:

- Dependency theory: development outcomes are shaped by structural inequalities created by global economic systems that benefit the Global North at the expense of the South (Deji, O.F. 2012); and
- Decolonial theory: calls for re-centring Southern epistemologies, agency, and self-determination in knowledge production and policy (Gascoigne, J., *et al* 2025).

### ***Problems with Global North prescription to Global South include:***

1. **Undermining local agency** by treating Southern contexts as passive recipients rather than active agents (Baguios, A., *et al*, 2021);
2. **Reinforcing power inequalities** by replicating neo-colonial relations through discourse and practice (Downes, G., 2013);

---

<sup>2</sup> India is around 1.45 billion and China is about 1.41 billion

3. **Marginalising Southern knowledge** by privileging Northern epistemologies and marginalising other ways of knowing (Kamruzzaman. P., 2024);
4. **Producing mis-aligned policies** by generating recommendations that do not fit the socio-political reality on the ground (Pritchett, L., *et al*, 2012); and
5. **Perpetuating stereotypes via framing** so simplistic global binaries demean or mis-represent (Teixeira da Silva, J., 2021).

## 3 Some Case Studies

There are some well-established, frequently cited case studies from education, health, and climate policy that empirically illustrate the problems created when Global North actors adopt a prescriptive, “telling the Global South what to do” tone. The cases below are drawn from peer-reviewed research, programme evaluations, and academic critical development. Each example is structured to show: (a) what was prescribed; (b) what went wrong; and (c) what the evidence shows about why tone mattered.

### 3.1 Education: Universal Schooling Models & Curriculum Transfer, Sub-Saharan Africa

**Case:** Driven Education Reforms (1990s–2000s)

**Where:** Multiple countries in Sub-Saharan Africa

**Who:** World Bank, bilateral donors, Northern NGOs

**Policy Tone:** “Best practice” schooling models and curricula should be adopted to achieve development and competitiveness

#### What was prescribed

- Universal primary education measured by enrolment numbers;
- Standardised curricula modelled on Western schooling systems;
- English/French language dominance; and
- Teacher accountability systems imported from the Global North.

#### What went wrong

- Enrolment increased, though learning outcomes stagnated or declined;
- Curricula were poorly aligned with local livelihoods, languages, and social realities;
- Teachers and communities felt *policed*, not supported; and
- Local knowledge systems were excluded from what counted as “education”

#### Key empirical findings

- Studies showed that education reforms prioritised *donor metrics* over locally-defined educational value;
- Parents and students often disengaged because schooling was perceived as externally imposed and culturally alien; and

- Learning crises persisted despite decades of reform.

### Key references

- Samoff, J. (2004): *From funding projects to supporting sectors*
- Tikly, L. & Barrett, A. (2011): *Education quality and postcolonial critique*
- Pritchett, L. (2013): *The Rebirth of Education*

### Why tone mattered

The prescriptive framing positioned Global North actors as *knowers* and Southern institutions as *implementers*, undermining local pedagogical authority and relevance.

## 3.2 Health: Structural Adjustment & Health System Reform

**Case:** IMF/World Bank Structural Adjustment Programmes

**Where:** Africa, Latin America, parts of Asia

**Policy Tone:** Countries *must* liberalise, privatise, and reduce public spending to achieve health efficiency

### What was prescribed

- Cuts to public health budgets;
- Introduction of user fees;
- Market-based health provision; and
- “Efficiency” framed through Northern economic models.

### What went wrong

- Sharp declines in access to healthcare for the poorest;
- Increased maternal and child mortality in several countries;
- Health systems weakened, not strengthened; and
- Long-term erosion of trust in public health institutions.

### Key empirical findings

- User fees dramatically reduced healthcare use among low-income populations;
- Countries that resisted or later reversed these prescriptions (e.g., Ghana and Uganda) saw improvements in health outcomes; and
- The “technical” economic advice ignored social realities and political constraints.

### Key references

- Cornia, G.A *et al* (1987): *Adjustment with a Human Face*
- Mkandawire, T. & C. Soludo (1999): *Our Continent, Our Future*
- Stiglitz, J. E. (2004): *Globalization and Its Discontents*

### Why tone mattered

Health reforms were framed as *non-negotiable technical necessities*, silencing local health professionals and policymakers who warned of social harm.

### 3.3 Health: Global Health Research & “Helicopter Science”

**Case:** HIV/AIDS and Global Health Trials

**Where:** Sub-Saharan Africa

**Who:** Northern universities, pharmaceutical firms

**Policy Tone:** Africa as a “testing ground” for global solutions

#### What was prescribed

- Research agendas set in the Global North;
- Limited local authorship or leadership; and
- Short-term data extraction with little capacity building.

#### What went wrong

- Ethical controversies over consent and standards of care;
- Local researchers marginalised;
- Interventions poorly adapted to local health systems; and
- Communities viewed research as exploitative.

#### Key references

- Crane, J. (2010): *Adverse events and global health research*
- Abimbola, S. (2019): *The foreign gaze*
- Costello, A. & A. Zumla (2000): *Moving beyond extractive research*

#### Why tone mattered

The implicit message was that knowledge flows one way, from North to South, reinforcing dependency and mistrust.

### 3.4 Climate Policy: REDD+ Forest Carbon Schemes

**Case:** Reducing Emissions from Deforestation and Forest Degradation)

**Where:** Brazil, Central Africa and Indonesia

**Who:** UN agencies, Northern governments, conservation NGOs

**Tone:** Southern countries *should* conserve forests to offset Northern emissions

#### What was prescribed

- Carbon accounting frameworks designed in the Global North;
- Forest protection enforced through external monitoring; and
- Market-based climate solutions.

#### What went wrong

- Indigenous land rights overridden or ignored;
- Local forest management practices delegitimised;
- Benefits flowed disproportionately to intermediaries; and
- Communities bore costs while Global North continued high emissions.

### Key empirical findings

- REDD+ often failed to reduce deforestation sustainably.
- Where it worked, it was locally designed and community-led, not imposed.
- Projects reproduced colonial conservation models ("fortress conservation").

### Key references

- Corbera, E. & H. Schroeder (2011): *REDD+ and power*
- Eberle, C. *et al* (2019): *Carbon Colonialism*
- Fairhead, J. *et al* (2012): *Green Grabbing: a new appropriation of nature*

### Why tone mattered

Climate responsibility was shifted southwards, with Northern actors positioning themselves as *designers and overseers*, not co-learners.

## 3.5 Climate & Development: "Good Governance" Conditionality

**Case:** Governance Conditionality in Aid & Climate Finance

**Where:** Africa, South Asia

**Tone:** Countries must adopt "good governance" frameworks to be trusted with funds

### What was prescribed

- Institutional reforms based on Western state models;
- Anti-corruption frameworks detached from political realities; and
- Uniform benchmarks of "good governance".

### What went wrong

- Reforms often became symbolic rather than functional;
- Local political settlements destabilised; and
- Governments focused on satisfying donors rather than citizens.

### Key references

- Andrews, M. *et al* (2017): *Building State Capability*
- Booth, D. (2012): *Working with the Grain*
- Grindle, M. (2004): *Good enough governance*

### Why tone mattered

Prescriptive governance language ignored that states develop through context-specific political processes, not checklists.

## 4 Cross-Case Patterns

Across sectors, the empirical record shows that prescriptive Global North tones tend to:

- Crowd out local knowledge rather than improve outcomes;
- Reproduce colonial power relations, even when intentions are good;
- Reduce effectiveness because solutions lack contextual fit; and
- Damage legitimacy and trust, undermining long-term sustainability.

Conversely, co-produced, locally-led approaches consistently perform better.

There are clear, increasingly well-documented examples of a *Global North-centric, prescriptive attitude* emerging in the advanced AI risk/AI safety discourse, and these examples are now being explicitly criticised by academics and policymakers from the Global South. While the field is newer than health or development, the patterns are strikingly similar. Below are clear examples, not just abstract critique, showing *how* this attitude appears, *why it is problematic*, and *what the evidence is so far*.

### 4.1 Frontier AI Risk Framed as a Universal Priority

**Case:** Existential and Frontier AI Risk Agendas (US/UK-led)

**Who sets the agenda:**

- US and UK governments;
- Elite universities (Oxford, Stanford, MIT);
- Frontier labs (OpenAI, Anthropic, DeepMind); and
- Northern philanthropic funders.

**Prescriptive message (implicit or explicit)**

All countries should prioritise frontier AI risk (AGI, catastrophic misuse, extinction-level events) and align with emerging global safety frameworks.

**Evidence of the problem**

- Surveys and consultation processes for global AI governance show low participation from low- and middle-income countries;
- Global South policymakers repeatedly state that immediate harms (labour displacement, surveillance, language exclusion, data extraction) are far more salient than speculative existential risk; yet
- Frontier risk dominates: global summits, safety frameworks, funding priorities and research prestige

**Key references**

- Abimbola, S. *et al* (2024): “Whose problems count?”
- Birhane, A & J. van Dijk (2020): *AI ethics framed through Western philosophical lenses*

Global South academics argue this creates agenda displacement: scarce governance capacity is re-directed to problems defined elsewhere.

### **Why tone matters**

The framing suggests that Global South countries are *behind* or *irresponsible* if they focus on near-term harms rather than speculative long-term risks, mirroring earlier development paternalism.

## 4.2 Compute, Capability, and “Responsible Scaling” Frameworks

**Case:** Responsible Scaling Policies and Compute Governance

### **What is being proposed**

- Caps or controls on frontier compute;
- Export controls on advanced chips; and
- Safety thresholds defined by Northern laboratories.

### **Empirical concern**

These frameworks are being designed without meaningful Global South input, yet would:

- Lock in existing technological hierarchies;
- Prevent late-industrialising countries from developing frontier capabilities; and
- Cement dependency on Northern firms.

### **Evidence from policy analysis**

Academics from Brazil, Africa and India have explicitly likened AI compute governance to:

- Nuclear non-proliferation regimes; and
- Climate mitigation rules that froze unequal development trajectories.

### **Key references**

- Abiade, S. F. (2025): *Algorithmic Sovereignty and the New Security Dependencies*
- Basu, S. & A. Omotubora (2024): *Decoding and Reimagining AI Governance*
- Srinivas, B. (2025): *Artificial Intelligence and the Future of Emerging Economies*
- Matthijs M Maas, M.M. (2025). *The Global AI Governance Architecture*
- Abungu, C., et al (2023): *Can apparent bystanders distinctively shape an outcome?*

### **Why tone matters**

The rhetoric often assumes:

*“We already have the capability; others must be constrained for global safety.”*

This echoes colonial conservation and “good governance” conditionality; safety for some at the cost of sovereignty for others.

## 4.3 AI Ethics & Safety Research as “Helicopter Science”

**Case:** AI Risk Research on Global South Contexts

### What happens

Northern researchers study:

- AI misuse risks in fragile states;
- Misinformation in African or Asian contexts; and
- Surveillance or authoritarian deployment risks.

### But

- Research questions are set in the North;
- Local academics are under-credited or absent; and
- Findings are used to justify external governance prescriptions.

### Empirical parallels

This has been explicitly compared to:

- Global health “helicopter research”<sup>3</sup>; and
- Development economics field experiments.

### Key references

- Birhane, A. (2021): *Algorithmic injustice: a relational ethics approach*
- Abimbola, S. (2019): *The foreign gaze*

### Why tone matters

Global South contexts are framed primarily as risk vectors, not sites of legitimate governance innovation or knowledge.

## 4.4 “If You Don’t Regulate, You’re a Risk” Framing

**Case:** AI Regulation Capacity Narratives

### Observed rhetoric

Claims that countries without strong AI regulation are:

- “Weak links” Roberts, H. *et al* (2024);
- “Backdoors for misuse” Latif, L. (2025); and/or
- “Global security risks” Amodei, D. (2025)

### Why this is empirically problematic

- Many Global South states lack regulatory capacity because of resource constraints and competing development priorities;
- The framing shifts blame *onto them*, rather than onto concentration of AI power in Northern firms; and
- Unequal access to compute and capital.

---

<sup>3</sup> refers to a practice where researchers from wealthier countries conduct studies in lower-income countries with little involvement of local researchers or community members

**This mirrors:**

- Pandemic preparedness narratives; and
- Financial regulation conditionality.

**Why tone matters**

Again, Global South contexts are framed primarily as risk vectors, not sites of legitimate governance innovation or knowledge.

## 5 Sector Cross-Cutting

Across these examples, the same empirical patterns appear as in education, health, and climate policy:

| Pattern                             | AI Risk Equivalent                          |
|-------------------------------------|---------------------------------------------|
| Agenda set in the North             | Frontier risk dominates                     |
| Universal prescriptions             | "Global alignment" rhetoric                 |
| Marginalisation of local priorities | Near-term harms sidelined                   |
| Power asymmetry                     | Compute, data, capital concentrated         |
| Moralising tone                     | "Responsible" versus "irresponsible" states |

**What Global South academics are explicitly saying**

Increasingly, the critique is *not* anti-AI safety, but anti-paternalistic AI safety:

- *"AI safety cannot become the new development conditionality"* Abungu, C., *et al* (2023); and
- *"Global AI governance risks reproducing digital colonialism under the banner of existential risk"* Gwagwa, A. (2024).

**Why this matters for Advanced AI Risk**

Ironically, the prescriptive tone may increase risk by:

- Reducing trust in global coordination;
- Encouraging non-alignment or quiet resistance;
- Pushing AI development into less cooperative regimes; and
- Undermining legitimacy of safety institutions.

This is exactly the failure mode documented in climate and health governance. Yet there are alternative governance models that avoid this North-South dynamic. There are alternative AI governance models that are explicitly designed (or could be designed) to avoid Global North - Global South paternalism, drawing on empirical precedents from climate, health, and technology governance *and* emerging AI-specific proposals.

## 6 Some Alternatives

### 6.1 Co-Production & Polycentric Governance

Instead of global alignment, there could be co-production and polycentric governance.

#### **Core idea**

AI risk governance should be co-produced across multiple centres of authority, not harmonised top-down around Northern risk framings.

#### **What this looks like in practice**

- Multiple regional AI risk forums (Africa, Latin America, South Asia, Southeast Asia);
- Each forum defines priority risks, acceptable trade-offs and local red lines; and
- Global coordination occurs *between* these centres, not *over* them.

#### **Precedents**

- Polycentric climate governance (Ostrom, E., 2010);
- Global health networks where regions define surveillance priorities (Africa CDC, 2024); and
- Internet governance via ICANN's multi-stakeholder model (imperfect, though instructive) (ICANN (2025)).

#### **Why this avoids the North-South problem**

- No single epistemic centre dictates what "counts" as AI risk;
- Frontier AI risk becomes *one input*, not the organising principle; and
- Global South actors are agenda-setters, not consultees.

#### **Trade-offs**

- Slower consensus;
- Harder to coordinate catastrophic-risk responses; and
- Requires trust in regional capacity.

### 6.2 Capability-Based Risk Governance

Similarly, instead of development-free risk thresholds, there could be capability-based risk governance.

#### **Core idea**

AI governance obligations should be proportional to actual AI capability and power, not formal state status.

#### **What this looks like**

- Regulatory burden tied to compute scale, model generality and deployment reach;
- Frontier labs (wherever located) face the highest obligations; and
- States with limited AI capacity are not forced into premature compliance.

### **Precedents**

- Common but Differentiated Responsibilities in climate law (United Nations Framework Convention on Climate Change, 1992)
- Asymmetric obligations in financial regulation (Basel Committee on Banking Supervision, IOSCO & IAIS (2010); and
- Public health surveillance thresholds UNHCR (2025).

### **Why this avoids the North–South problem**

- Shifts responsibility away from poorer states;
- Prevents AI safety becoming a new development conditionality; and
- Recognises that *risk creation is highly concentrated*.

### **Trade-offs**

- Politically difficult for powerful states and firms;
- Requires credible metrics of AI capability; and
- Risk of regulatory arbitrage if poorly designed.

## **6.3 Risk Pluralism**

Instead of a single-narrative AI safety, there could be risk pluralism.

### **Core idea**

Accept that AI risk is not a single objective phenomenon, but a plural one shaped by context.

### **What this looks like**

- Explicit recognition of multiple legitimate risk categories such as: existential/catastrophic, economic and labour, political and surveillance, and linguistic and cultural; and
- Governance institutions must justify why one risk is prioritised over another.

### **Precedents**

- Health priority-setting frameworks;
- Environmental justice movements; and
- Participatory technology assessment (Brazil and Denmark).

### **Why this avoids the North–South problem**

- Prevents Northern speculative risks from crowding out Southern-lived harms;
- Makes risk prioritisation a political choice, not a technical inevitability; and
- Legitimises disagreement.

### **Trade-offs**

- Less elegant global messaging;
- Harder to mobilise around a single narrative; and
- Requires institutional humility.

## 6.4 Infrastructure Justice & Access-First Governance

Without equitable access to AI infrastructure, safety governance risks consolidating power and entrenching hierarchical structures. In other words, AI safety governance, when divorced from equitable infrastructure access, becomes a mechanism for preserving hierarchy rather than mitigating risk.

### Core idea

AI safety governance must be coupled with equitable access to AI infrastructure, or it entrenches hierarchy.

### What this looks like

- Shared public compute infrastructure (regional AI clouds);
- Open models and multilingual datasets;
- Technology transfer as part of safety agreements; and
- Funding for local model development.

### Precedents

- Green Climate Fund (imperfect, though relevant);
- Open-source vaccine platforms; and
- Satellite data commons.

### Why this avoids the North-South problem

- Prevents safety rules from freezing existing technological dominance;
- Recognises that *capacity inequality itself is a risk*; and
- Reframes safety as empowerment, not restraint.

### Trade-offs

- Expensive;
- Raises proliferation concerns; and
- Requires trust-building and safeguards.

## 6.5 South-Led Norm Entrepreneurship

There could be South-led norm entrepreneurship instead of Northern moral leadership.

### Core idea

Global South states and institutions act as norm creators, not norm takers.

### What this looks like

- Regional AI governance frameworks that precede global ones, export norms upward; and
- Examples: African Union AI Strategy, Latin American digital rights frameworks, and the India public digital infrastructure model.

### Precedents

- TRIPS flexibilities in global IP law;
- Right-to-development norms; and

- Data sovereignty movements.

#### **Why this avoids the North-South problem**

- Reverses epistemic flow;
- Forces Global North actors to adapt; and
- Builds legitimacy through ownership.

#### **Trade-offs**

- Fragmentation risk
- Limited enforcement power initially; and
- Requires regional coordination capacity.

## **6.6 Procedural Justice in Global AI Institutions**

People accept decisions when they trust the process behind them, not just when they agree with the result.

#### **Core idea**

Legitimacy comes from how decisions are made, not just their content.

#### **What this looks like**

- Co-chairing of global AI bodies by North and South;
- Rotating secretariats;
- Mandatory representation quotas; and
- Budgetary independence from Northern donors.

#### **Precedents**

- WHO governance reforms;
- UN climate negotiations; and
- Some multilateral development banks.

#### **Why this avoids the North-South problem**

- Reduces agenda capture;
- Makes paternalism structurally harder; and
- Builds long-term trust.

#### **Trade-offs**

- Bureaucratic complexity;
- Slower decision-making; and
- Political negotiation overhead.

## **6.7 Reflexive Safety**

Reflexive safety treats risk, harm, and responsibility as evolving rather than settled. Instead of assuming that safety can be guaranteed through fixed rules or moral certainty about what is “right,” it emphasises continual reflection, learning, and adjustment as systems, contexts, and impacts change.

Reflexive safety acknowledges uncertainty, power asymmetries, and the limits of foresight, and therefore prioritises feedback, contestability, and the capacity to revise decisions over time.

By contrast, moral certainty tends to frame safety as a matter of enforcing predefined values or outcomes, often insulating decision-makers from challenge. While this can produce clarity and speed, it risks rigidity: locking in assumptions, excluding dissenting perspectives, and failing to detect emergent harms. Reflexive safety resists this closure. It shifts the focus from being *right once* to remaining *responsive over time*.

In practice, reflexive safety means designing governance and technical systems that expect disagreement, enable critique, and institutionalize mechanisms for correction, such as ongoing monitoring, participatory review, and the ability to pause, adapt, or roll back decisions. Safety, on this view, is not a final state to be achieved, but a continual process of reflection and recalibration in the face of uncertainty.

### **Core idea**

AI safety institutions should be explicitly reflexive about power, not morally authoritative.

### **What this looks like**

- Regular review of whose risks are prioritised;
- Sunset clauses on safety frameworks;
- Independent audits of epistemic bias; and
- Acknowledgement of uncertainty and disagreement.

### **Precedents**

- Responsible innovation frameworks;
- Technology assessment bodies; and
- Feminist and decolonial research ethics.

### **Why this avoids the North-South problem**

- Undercuts the “we know best” posture;
- Normalises contestation; and
- Encourages learning rather than compliance.

### **Trade-offs**

- Less rhetorical clarity;
- Harder to justify emergency powers; and
- Requires cultural change in AI safety community.

## 7. Comparisons of Different Alternatives

Below is a comparative summary table of the different alternatives.

| Model                   | Avoids Paternalism By       | Main Risk            |
|-------------------------|-----------------------------|----------------------|
| Polycentric governance  | Decentralising authority    | Fragmentation        |
| Capability-based duties | Targeting real risk creator | Political resistance |
| Risk pluralism          | Legitimate disagreement     | Loss of focus        |
| Infrastructure justice  | Equalising capacity         | Proliferation fears  |
| South-led norms         | Reversing epistemic flow    | Limited reach        |
| Procedural justice      | Fair decision-making        | Bureaucracy          |
| Reflexive safety        | Humility about knowledge    | Slower action        |

Avoiding the North–South dynamic in AI risk governance is not about rejecting safety. It is about rejecting the *assumption* that:

*“There is one correct framing of AI risk, and it happens to be ours.”*

The strongest governance models distribute authority, target power where it actually lies, treat disagreement as legitimate, and couple restraint with empowerment.

## 8. References

Abiade, S. F. (2025). Algorithmic Sovereignty and the New Security Dependencies: How Foreign AI Surveillance Technologies Reshape Domestic Autonomy in the Global South. *World Journal of Advanced Research and Reviews*, Vol 27 (2) pp.162-180.  
[10.30574/wjarr.2025.27.2.2845](https://doi.org/10.30574/wjarr.2025.27.2.2845)

Abimbola, S. (2019). The foreign gaze: authorship in academic global health. *BMJ Global Health*. <https://doi.org/10.1136/bmjgh-2019-002068>

Abimbola, S., van de Kamp, J., Lariat, J., Rathod, L., Klipstein-Grobusch, K., van der Graaf, R., & Bhakuni, H. (2024). Unfair knowledge practices in global health: a realist synthesis. *Health Policy and Planning*, Vol. 39, (6) pp. 636–650.  
<https://doi.org/10.1093/heapol/czae030>

Abungu, C., Malonza, M. & S.N. Adan (2023). Can apparent bystanders distinctively shape an outcome? Global south countries and global catastrophic risk-focused governance of artificial intelligence. [arXiv:2312.04616](https://arxiv.org/abs/2312.04616)

Africa CDC (2024) Strengthening cross-border surveillance and information sharing in Africa: Strategic framework.  
<https://africacdc.org/wp-content/uploads/2024/10/Strengthening-Cross-Border-Surveillance-and-Information-Sharing-in-Africa-strategic-framework.pdf>

African Union (2024). Continental Artificial Intelligence Strategy (2024). African Union  
[https://au.int/sites/default/files/documents/44004-doc-EN-Continental\\_AI\\_Strategy\\_July\\_2024.pdf](https://au.int/sites/default/files/documents/44004-doc-EN-Continental_AI_Strategy_July_2024.pdf)

Adan, S. N. & J. Wiaterek (2025). AI Benefit-Sharing Framework: Balancing Access and Safety. AI Governance Initiative, Oxford Martin School, University of Oxford.  
<https://aigi.ox.ac.uk/publications/ai-benefit-sharing-framework-balancing-access-and-safety/>

Andrews, M., Pritchett, L. & M. Woolcock (2017). Building State Capability: Evidence, Analysis, Action. Oxford online, Oxford Academic.  
<https://doi.org/10.1093/acprof:oso/9780198747482.001.0001>

Amodei, D. (2025) Statement from Dario Amodei on the Paris AI Action Summit, Anthropic  
<https://www.anthropic.com/news/paris-ai-summit>

Baguios, A., King, M., Martins, A. and Pinnington, R. (2021) Are we there yet? Localisation as the journey towards locally led practice: models, approaches and challenges. London: Overseas Development Institute.  
<https://odi.org/en/publications/are-we-there-yet-localisation-as-the-journey-towards-locally-led-practice/>

Basel Committee on Banking Supervision, IOSCO & IAIS (2010) Review of the Differentiated Nature and Scope of Financial Regulation, Bank for International Settlements.  
<https://www.bis.org/publ/joint24.pdf>

Basu, S. & A. Omotubora (2024): Decoding and Reimagining AI Governance Beyond Colonial Shadows. Handbook on Public Policy and Artificial Intelligence, pp. 222-236.  
[10.4337/9781803922171.00025](https://doi.org/10.4337/9781803922171.00025)

Birhane, A., & J. van Dijk (2020). Robot Rights? Let's Talk about Human Welfare Instead Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society.  
[arXiv:2001.05046](https://arxiv.org/abs/2001.05046)

Birhane, A. (2021). Algorithmic injustice: a relational ethics approach', Patterns, 2 (2),  
<https://doi.org/10.1016/j.patter.2021.100205>

Booth, D. (2012). Working with the Grain and Swimming against the Tide. Public Management Review, Vol 14.  
[https://www.researchgate.net/publication/254349646\\_Working\\_with\\_the\\_Grain\\_and\\_Swimming\\_against\\_the\\_Tide/citation/download](https://www.researchgate.net/publication/254349646_Working_with_the_Grain_and_Swimming_against_the_Tide/citation/download)

Campbell-Stephens, R.M. (2021) Educational Leadership and the Global Majority - Decolonising Narratives. Springer Nature Switzerland.  
<https://link.springer.com/book/10.1007/978-3-030-88282-2>

Cornia, G. A., Jolly, R., & Stewart, F. (1987). Adjustment with a human face. UNICEF  
<https://digitallibrary.un.org/record/46296?ln=en>

Corbera, E. & H. Schroeder (2011). Governing and implementing REDD+. Environmental Science & Policy, Vol. 14, (2) pp. 89-99. <https://doi.org/10.1016/j.envsci.2010.11.002>

Costello, A. and Zumla, A. 2000. Moving to research partnerships in developing countries. British Medical Journal, Vol. 321, pp. 827-29.  
<https://pmc.ncbi.nlm.nih.gov/articles/PMC1118627/>

Crane, J. (2010). Adverse events and placebo effects: African scientists, HIV, and ethics in the 'global health sciences'. Social Studies of Science, Vol 40 (6) 843-870.  
[https://www.researchgate.net/publication/51108816\\_Adverse\\_events\\_and\\_placebo\\_effects\\_African\\_scientists\\_HIV\\_and\\_ethics\\_in\\_the\\_'global\\_health\\_sciences'](https://www.researchgate.net/publication/51108816_Adverse_events_and_placebo_effects_African_scientists_HIV_and_ethics_in_the_'global_health_sciences')

Deji, O.F., 2011. *Gender and rural development: Advanced studies* (Vol. 2). LIT Verlag Münster.  
[https://scholar.google.com/citations?view\\_op=view\\_citation&hl=en&user=i3ZRZDEAAAAJ&citation\\_for\\_view=i3ZRZDEAAAAJ:7PzIFSSx8tAC](https://scholar.google.com/citations?view_op=view_citation&hl=en&user=i3ZRZDEAAAAJ&citation_for_view=i3ZRZDEAAAAJ:7PzIFSSx8tAC)

Downes, G (2013) 'A Critical Analysis of North-South Educational Partnerships in Development Contexts', Policy and Practice: A Development Education Review, Vol. 16, pp. 1-12.  
<https://www.developmenteducationreview.com/issue/issue-16/critical-analysis-north-south-educational-partnerships-development-contexts>

- Eberle, C., Münstermann, N. & J. Siebeneck (2019). Carbon Colonialism: A postcolonial assessment of carbon offsetting.  
[https://www.researchgate.net/publication/337622634\\_Carbon\\_Colonialism\\_A\\_postcolonial\\_assessment\\_of\\_carbon\\_offsetting](https://www.researchgate.net/publication/337622634_Carbon_Colonialism_A_postcolonial_assessment_of_carbon_offsetting)
- Fairhead, J., Leach, M., & I. Scoones, I. (2012). Green Grabbing: a new appropriation of nature? *The Journal of Peasant Studies*, 39 (2) pp 237–261.  
<https://doi.org/10.1080/03066150.2012.671770>
- Gascoigne, J., *et al* (2025). 'Decolonising Research: Tackling obstacles and working through challenges' Interdisciplinary Global Development Centre Working Paper No. 6, University of York. <https://doi.org/10.15124/j2d6-d689>
- Gerber, J-F., *et al* (2018). "Post-Growth in the Global South? Some Reflections from India and Bhutan". *Ecological Economics*. 150: 353–358.  
<https://www.sciencedirect.com/science/article/abs/pii/S0921800916315567>
- Grindle, M. (2004). Good Enough Governance: Poverty Reduction and Reform in Developing Countries. *Governance*, Vol 17 (4) pp 525-548.  
<https://doi.org/10.1111/j.0952-1895.2004.00256.x>
- Gwagwa, A. (2024) Resisting colonialism - why AI systems must embed the values of the historically oppressed, Chatham House  
<https://www.chathamhouse.org/2024/06/artificial-intelligence-and-challenge-global-governance/06-resisting-colonialism-why-ai>
- ICANN (2025) In Brief: Multistakeholder Internet Governance, Internet Corporation for Assigned Names and Numbers.  
<https://www.icann.org/en/system/files/files/in-brief-multistakeholder-internet-governance-28oct25-en.pdf>
- Kamruzzaman, P. (2024). Exploring challenges and politics of knowledge production in the global South - Evidence from Bangladesh's aid and development sector. Volume 43, Issue 5.  
<https://doi.org/10.1177/23996544241299931>
- Latif, L. (2025). Regulating Digital Health and Health Apps in South Africa: Lessons Learnt, SSRN Electronic Journal <https://ssrn.com/abstract=5122390>
- Leal, J., *et al* (2022). Decolonising ecological research: a debate between global North geographers and global South field ecologists.  
[https://www.researchgate.net/publication/364265422\\_Decolonising\\_ecological\\_research\\_a\\_debate\\_between\\_global\\_North\\_geographers\\_and\\_global\\_South\\_field\\_ecologists](https://www.researchgate.net/publication/364265422_Decolonising_ecological_research_a_debate_between_global_North_geographers_and_global_South_field_ecologists)
- Matthijs M Maas, M. M. (2025). The Global AI Governance Architecture: Past and Futures. Oxford Academic. <https://doi.org/10.1093/9780191988455.003.0005>

Mkandawire, T. & C. Soludo (1998) African Perspectives on Structural Adjustment. Council for the Development of Social Science Research in Africa.

<https://idrc-crdi.ca/sites/default/files/openbooks/855-4/index.html>

Ostrom, E. (2010). Polycentric systems for coping with collective action and global environmental change, *Global Environmental Change*, 20 (4) pp. 550–557.

<https://doi.org/10.1016/j.gloenvcha.2010.07.004>

Pritchett, L., *et al* (2012) *Looking like a state: techniques of persistent failure in state capability for implementation*. CID Working Paper No. 239. Cambridge, MA: Harvard Kennedy School, Center for International Development.

<https://www.hks.harvard.edu/centers/cid/publications/faculty-working-papers>

Pritchett, L. (2013) *The Rebirth of Education: Schooling Ain't Learning*, Center for Global Development, The Harvard Center for International Development.

<https://www.hks.harvard.edu/centers/cid/publications/books/the-rebirth-of-education-schooling-aint-learning>

Roberts, H., Hine, E., Taddeo, M. & L. Floridi (2024) Global AI governance: barriers and pathways forward, *International Affairs*, 100 (3).

[https://www.researchgate.net/publication/379662810\\_Global\\_AI\\_governance\\_barriers\\_and\\_pathways\\_forward](https://www.researchgate.net/publication/379662810_Global_AI_governance_barriers_and_pathways_forward)

Roche, C. *et al* (2021). Artificial Intelligence Ethics: An Inclusive Global Discourse? In proceedings of the 1<sup>st</sup> Virtual Conference on Implications of Information and Digital Technologies for Development. <https://doi.org/10.48550/arXiv.2108.09959>

Rodríguez, A. L. G. (2022) *Development Studies from a Decolonial Perspective: Discourse Analysis on the OECD Development Reports*. Bachelor Thesis in Peace & Development Studies. Manuela Nilsson University.

<https://efaidnbmnnnibpcajpcglclefindmkaj/https://www.diva-portal.org/smash/get/diva2:1751061/FULLTEXT01.pdf>

Samoff, J. (2004). Forthcoming. "From Funding Projects to Supporting Sectors? Observations on the Aid Relationship in Burkina Faso." *International Journal of Educational Development*.

<https://scholar.google.com/scholar?q=Samoff,+Joel.+Forthcoming.+%E2%80%9CFrom+Funding+Projects+to+Supporting+Sectors?+Observations+on+the+Aid+Relationship+in+Burkina+Faso.%E2%80%9D+International+Journal+of+Educational+Development>

Srinivas, B. (2025): Artificial Intelligence and the Future of Emerging Economies. *International Journal of Multidisciplinary research*, Vol 3 (7)

<https://doi.org/10.5281/zenodo.16791565>

Stiglitz, J. E. (2004). *Globalization and Its Discontents*. *Public Choice* Vol. 120, No. 1/2, pp. 234-239. <https://www.jstor.org/stable/30025840>

Teixeira da Silva, J. (2021). Rethinking the use of the term 'Global South' in academic publishing. European Science Editing.  
[https://www.researchgate.net/publication/354152094\\_Rethinking\\_the\\_use\\_of\\_the\\_term\\_'Global\\_South'\\_in\\_academic\\_publishing](https://www.researchgate.net/publication/354152094_Rethinking_the_use_of_the_term_'Global_South'_in_academic_publishing)

Tikly, L. & Barrett, A. (2011). Social justice capabilities and quality of education in low-income countries. International Journal of Educational Development, Vol 31.  
[https://www.researchgate.net/publication/253415007\\_Social\\_justice\\_capabilities\\_and\\_quality\\_of\\_education\\_in\\_low\\_income\\_countries](https://www.researchgate.net/publication/253415007_Social_justice_capabilities_and_quality_of_education_in_low_income_countries)

United Nations Framework Convention on Climate Change (1992). *Article 3(1)*, adopted at the United Nations Conference on Environment and Development (Rio Earth Summit).  
<https://unfccc.int/resource/ccsites/zimbab/conven/text/art03.htm>

UNHCR (2025) Disease surveillance thresholds, UNHCR Emergency Health & Nutrition.  
<https://emergency.unhcr.org/emergency-assistance/health-and-nutrition/disease-surveillance-thresholds>

UNICEF (2025). Artificial Intelligence Governance in Motion: A rapid review of AI regulation and its implications for children's rights.  
<https://www.unicef.org/documents/artificial-intelligence-governance-motion>

Wells, T. R. (2025). Extending moral criticism beyond the Global North. *Journal of Global Ethics*, 21(1), 65–76. <https://doi.org/10.1080/17449626.2025.2506631>  
World Bank (2024) World Population Open Data.  
<https://data.worldbank.org/?locations=CN-IN-RU-BR-ZA>

This is a report by Global AI Governance Alliance with support and co-sponsorship of the **European Center for Peace and Development – University for Peace established by the United Nations.**

**Global AI Governance Alliance (GAIGANow)** is building a broader Alliance committed to advancing safe, effective, accountable and inclusive global AI governance to help steer AI progress towards benefiting humanity and supporting a more just world.

---

## Written By Richard Pagett

Full Professor for Sustainable Development  
European Center for Peace and Development  
UN University for Peace  
Belgrade, Serbia

## Suggested Citation

Pagett, R., 2026. Tone Matters  
Global AI Governance Alliance

## tone matters

Global AI Governance Alliance | [www.gaiganow.org](http://www.gaiganow.org)

Cover design, text and layout by GAIGANow  
Copyright 2026 © Global AI Governance Alliance