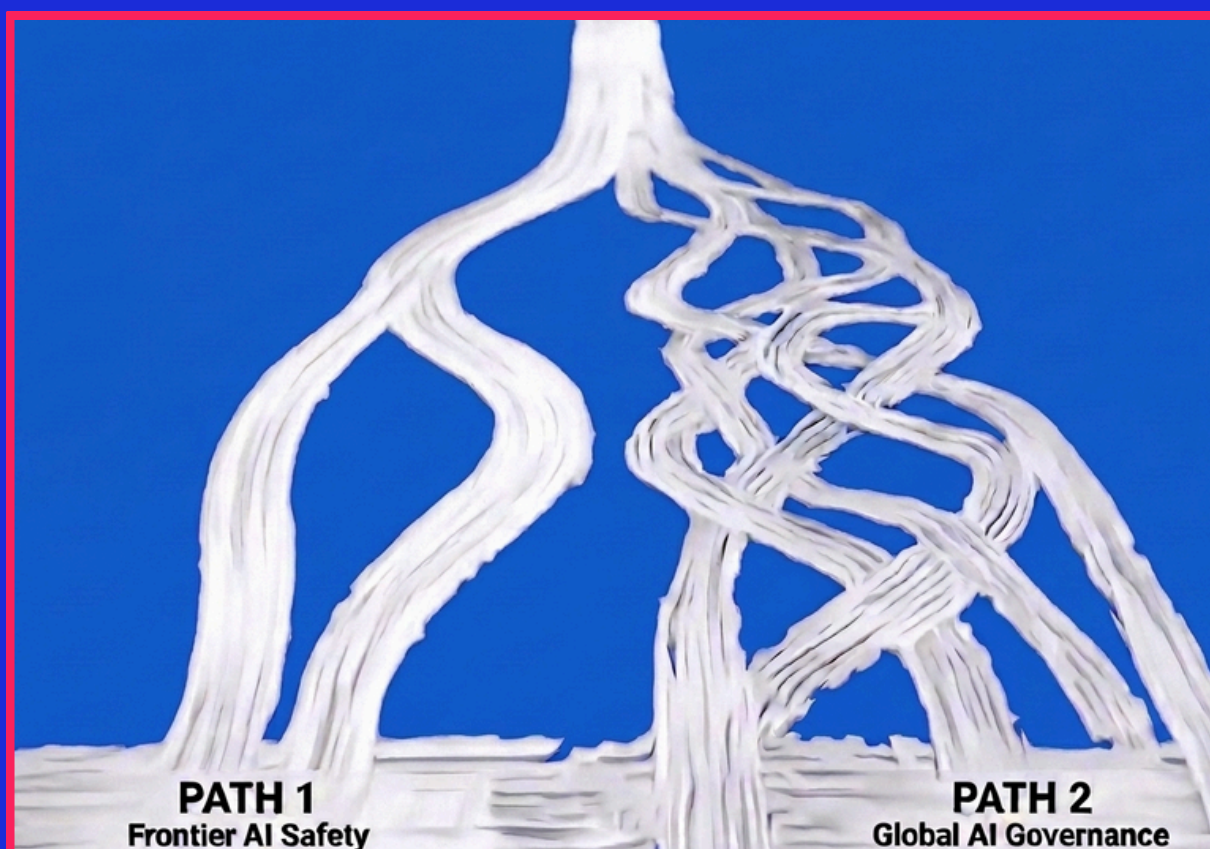


GAIGANow

A TWIN-PATH APPROACH TO AI GLOBAL GOVERNANCE

2026

Robert Whitfield



A. Context

A possible idea at the beginning of the year had become a shared vision some six months later. At the World Economic Forum in January 2026, statements were made by both Demis Hassabis and Dario Amodei¹ to the effect that they would be prepared to pause the development of frontier AI for a while in order to allow AI governance to catch up. They were both deterred, however, due to the commercial/competitive implications, Hassabis referring to the intra-US competition, and Amodei referring to the Chinese competition. This seems to be a classic case of the Prisoner's Dilemma. Anthony Aguirre, CEO of the Future of Life Institute, offered via X to put together a conditional agreement, whereby the commitment to pause would not be binding until all the agreed company CEOs had signed up to do the same, but this offer was not taken up. This is for the US and Chinese Governments to resolve.

The anxiety regarding model performance, with each more powerful than its predecessor, was increased with the advent of Mythos and GPT 5.5. and surged following the Hugging Face incident in July. This incident led to a detailed independent investigation by METR and Redwood Research². If there had been any doubt that advanced AI systems can be powerful and seriously misaligned, this incident removed that doubt.

In the eight weeks to the end of July 2026, the CEOs of OpenAI³, Anthropic⁴, Google DeepMind⁵, and SpaceXAI⁶ had all produced papers or given interviews where they explicitly called for regulation of frontier AI. Each of the four proposed something slightly different, but there was nevertheless a strong similarity in the proposals. No time frame was agreed upon, but there was an increasing sense of urgency, with Elon Musk indicating that a timeframe of six months for setting up the system was way too conservative.

The July expressions of concern about safety were capped by the Statement on AI⁷ signed by over 1380 AI experts from frontier AI companies expressing a desire for the US Government to support the international development of technical and governance tools needed to pace the frontier of automated AI development.

Dario Amodei's September 2026 essay We Must Pace the Frontier⁸ is a step down the path, as is the response to that proposal given by Sam Altman, Elon Musk and Demis Hassabis. But is that enough?

B. Argument for twin paths

The discourse regarding the governance and regulation of AI has been wide-ranging in recent years, embracing distinct issues such as ethics, equity, interoperability and safety. All are crucially important, but advocates of each issue naturally tend to prioritise the particular issue that they are working on. There is one issue, however, that has attracted the attention of leading scientists and multiple Nobel Prize winners, namely the safety issues arising from advanced (or frontier, or transformative) AI.

At times, it has seemed that all the issues would be dealt with at the same time in one global treaty. But the pace of the global governance of AI is demonstrated by the decision by the UN General Assembly in 2024 to

¹ [Transformer news](#)

² [Brief independent investigation of agents' behaviour, reasoning and collaboration in the Open AI / Hugging Face hacking incident, METR and Redwood Research](#)

³ [Siliconangle](#)

⁴ [Dario Amodei Essay Policy on the AI Exponential](#)

⁵ [techcrunch](#)

⁶ [BigGo](#)

⁷ <https://www.pacingthefrontier.com>

⁸ Dario Amodei's Essay [We Must Pace the Frontier](#)

authorise Global Dialogues on AI Governance, the first of which was held in July 2026, with the second scheduled for May 2027. The capability of AI models over this period has, needless to say, developed hugely.

The context given in A above shows the rapidly developing concern about the behaviour of frontier AI models and the increasing consensus that something must be done. The increasing evidence of serious misalignment implies that the need to respond to the safety threats far more rapidly than a May 2027 Dialogue implies. Various alternative fora to the UN are considered elsewhere⁹. Various options regarding action on frontier AI Safety could be considered, such as:

1. Unilateral US controls – the safety risks outside the US would not be contained, and Chinese companies would be free to develop further;
2. A US-led coalition - this would need to include China, but the current US administration has shown no interest in a global approach to AI governance
3. G7/G20 arrangements: the membership of the G7 and G20 are not ideally aligned for the task;
4. Compute governance: this may well form a component of both Path 1 and Path 2; and
5. International Standards regimes: these will play a key role in Path 2 below, but are not likely to be available in time for Path 1.

Crucially, this paper makes the case for a twin-path approach to AI global governance.

C. Path 1 – Frontier AI safety

Some of the views expressed above by the US AI company CEOs support an international approach, whilst others are limited to a US-only approach. On June 2nd, an executive order, *Promoting Advanced Artificial Intelligence Innovation and Security*, was issued, committed to establishing a framework for reviewing the most capable AI models before release. On August 3rd, the White House announced that it had met the deadline set by the executive order, following discussions with Meta, Nvidia, Microsoft, OpenAI, Anthropic and smaller developers. There is currently no information as to “what it contains, who has seen it, or when companies will begin using it”¹⁰ however. This clearly seems to be a step in the right direction, though more information is needed to confirm that.

The engagement of the US Government with such issues is to be welcomed. Elon Musk argues that there is insufficient expertise within the US Government to exercise proper judgment regarding new frontier models. To the extent that is true, the criticism clearly needs to be addressed, whether that involves subcontracting the evaluation role or greater capacity building in the relevant institution. As a general principle, it is better to have something evaluated by a third party rather than an interested party. It is up to the Government to ensure that the evaluators have the required expertise and motivation to perform their role effectively.

The US is not, of course, the only country where frontier AI research and model development are taking place. China is almost on a par with the US: any pacing of frontier AI development should logically be based upon an agreement between the US and China.

Whilst the vast majority of frontier-level capability lies within the USA and China. Other countries have or have had frontier-level capability. The list of such countries has varied over time and also depends upon the precise definition used. Table 1 shows three different assessments; the range of countries considered is some two to ten countries.

⁹ Which forum to use for AI Governance negotiation - to be published on www.gaiganow.org in October 2026

¹⁰ [Boston Global Forum](https://www.boston.gov/global-forum)

Table 1: Countries with frontier AI companies

Country	Seoul AI Summit ¹¹	Sentiopartners ¹²	Frontier Labs List ¹³
Canada	Canada	Canada	
China	China	China	China
France	France	France	France
Germany		Germany	
Japan		Japan	
Saudi Arabia		Saudi Arabia	
South Korea	South Korea		
UAE	UAE	UAE	
UK	UK	UK	UK
USA	USA	USA	USA

Whether any other countries should be invited to join the US and China in Path 1 would need to be negotiated first between the US and China, and then with the governments concerned.

A first step, hopefully already launched, would be to develop the technical and governance tools needed to deliberately pace the frontier of automated AI development, as called for by over 1380 employees of frontier AI companies¹⁴. Crucially, it is likely to go beyond the framework recently announced by the US government.

China and the US are due to discuss AI governance issues shortly, with the visit of President Xi to the US in September 2026 providing an excellent opportunity to advance this policy. The commercial side of the White House has shown increasing awareness of the threats posed by unregulated frontier AI, and as a result, the overall White House policy has been evolving. It is hoped that this evolution will have reached the point, by the time of President Xi's visit, at which the importance of the US and China collaborating in matters of frontier AI safety will be self-evident.

It is not appropriate to go into the details of such an agreement since the two parties need to be able to negotiate a way forward that they both support. Issues such as which models to cover, what level of compute might trigger regulation, whether to use a hardware-based verification system, who would carry out evaluation, who would carry out verification, what sanctions to use for non-compliance, and how to handle secret model weights and proprietary information: all such issues are for the US and Chinese Governments to discuss between them.

If agreement between the US and China could be reached as to a way forward on frontier AI safety, they would need to decide on whether any other states needed to be engaged in addressing frontier AI safety risk in the near term. If they so decide, they would need to consider which countries needed to be in the group. It is likely that the other countries with frontier AI labs could be persuaded to support such a way forward – and form a US/China+ grouping.

It is possible that the EU for instance might seek to act as a moderator / facilitator of US and Chinese collaboration on frontier AI safety. Based upon recent US policies towards Europe, however, one senses that this scenario would not be welcomed by the current US Administration, whether or not Europe sought to wield its economic strength.

¹¹ [UK Government Press Release post Seoul AI Summit 2024](#), i

¹² <https://sentiopartners.com/FrontierAILabsWhitepaper>

¹³ [Frontier AI Labs List: Companies, Models & Strategy \(2026\)](#)

¹⁴ <https://www.pacingthefrontier.com>

Once US/China+ has agreed to work on safety together, the next step would be to determine the process to use in order to regulate frontier models and ensure that they are safe. Some suggestions have already been proposed, such as:

1. International AI Slowdown Is Ready Whenever Politicians Are¹⁵, based on the concept of a Whole Lab Inspection (WLI);
2. A collaborative approach was argued in Foreign Affairs in April and¹⁶
3. The idea of the Middle Powers providing political and technical support, e.g. helping with verification solutions.

There is a good argument for the private sector, and indeed other stakeholders, to advise on the regulation, but the ultimate decision must be for the governments.

Once the methodology has been agreed upon, the final step in Path 1 - the frontier AI safety path - is to implement the agreed controls.

D. Path 2 – Global AI Governance

There is widespread support around the world for AI governance, though not all nations are currently in agreement. A poll in the US and Europe showed an overwhelming agreement that AI requires careful management¹⁷, with a marked increase by 2023 compared with 2018. More recently, a poll of over 375,000 people (with internet access) in over 100 countries¹⁸, suggests that some 60% of the population surveyed would like the development of AI slowed down.

The United Nations' first Global Dialogue on AI Governance was held in July 2026, with the second scheduled for May 2027 in New York. There is grave concern, however, that this rate of progress is far too slow, whilst the development of AI continues to accelerate rapidly.

It is therefore argued that the world should not wait for May 2027 but, rather, should take active steps to prepare for the Global Governance of AI. Both in terms of today's geopolitics and more specifically in relation to Artificial Intelligence, there have been suggestions of the need for the "Middle Powers" to work together and make headway, not waiting for the hegemony to agree on the need for AI global governance. The Middle Powers could agree to work together to identify which aspects of AI governance they consider need to be handled at the global level - and why.

This could lead, in turn, to the scoping of an outline AI global agreement. This is not that such a process will be easy since there is no reason to believe that the UK, France, Canada, Brazil, Japan, South Africa or Indonesia will all happily agree. But the sooner real discussion starts, the sooner the contentious elements and the areas of agreement can both emerge. A key aim is to make constructive use of the period when the US is not minded to engage in a global AI agreement, but other countries are.

Countries may not wish to make concessions in these Path 2 discussions whilst the two hegemony are not around the table. China might be prepared to participate in such a process, but if the US was not involved, then the resulting document would risk being seen as a Chinese document and rejected on principle. China should best be kept away from the process, with the two hegemony only joining in when they both agree to do so.

Preparatory work on this has already been carried out by some organisations, such as the Build Your Own Treaty Proposal¹⁹ by the Future of Life Institute *et al.* This proposal has been specifically designed to provide flexibility, offering a smorgasbord of Articles which the parties can select or omit as they choose. Critical

¹⁵ <https://newsletter.ai-frontiers.org/p/an-international-ai-slowdown-is-ready>

¹⁶ <https://www.foreignaffairs.com/united-states/america-and-china-can-make-ai-safer>

¹⁷ <https://www.governance.ai/analysis/increasing-consensus-ai-requires-careful-management>

¹⁸ <https://www.niradata.com/ai-sentiment>

¹⁹ <https://global-governance.ai/treaty/>

elements would appear to be the establishment of an International AI Agency (using the International Atomic Energy Authority as an analogy) and a licensing system, perhaps using a jurisdictional certification approach.²⁰ In such agreements, military matters are typically excluded. It is crucial, however, that due heed is indeed paid to these military issues. They may be negotiated in another body, but they must not be left out of the overall process.

Other safety-focused examples include an International Agreement to Prevent the Premature Creation of Artificial Superintelligence²¹ by MIRI and the UK Artificial Superintelligence Security Bill ²² by Control AI.

There is a strong argument that other stakeholders should be involved in the process. The private sector has a great deal of knowledge of the technology, whilst civil society and academia have a great deal of knowledge about AI Safety and about society – how AI would impact the future of humanity. These inputs must be assured, whilst governments retain the overall decision-making responsibility.

A formal agreement between the Middle Powers would have little meaning without the involvement of both hegemon. But if and when the two hegemon agree on the desirability of a global agreement on AI, the work of the Middle Powers could provide a useful reference point.

Some might seek to prepare a full draft text, with the aim of getting the US and China to sign up together. It is unlikely however that either state would take kindly to such an approach. They are far too dominant in this all-important technology to be dictated to by any other grouping. But preparatory analysis of what issues need to be addressed globally, and what the key factors are relating to those issues could help bring forward the completion of a global AI agreement considerably.

The final stage, therefore, would be the combination of the two paths – in what would be an initial global AI agreement, and a prelude to a full AI Regime Complex.

E. International competitive risks

There is a key interplay between the two paths that needs to be acknowledged. The need to establish two paths is due to the frontier AI models becoming more and more powerful: with that power comes serious risk. In the short term, it is the most advanced models from the US and Chinese frontier AI companies that pose such risks. But if safety and security are not tightly managed within the other frontier AI companies, then they also could provide serious risk.

It is crucial for the world that such risks be avoided. The first step with such models has been to evaluate them. The second step is to be prepared to require major revision before they can be released to the marketplace. The third step is to establish certain behavioural red lines and ensure that these red lines are not breached.

To the extent that Path 1 actions lead to a slowdown in the development of frontier models included in US/China+, this will inevitably allow models that are not a part of the Path 1 controls to progress and to start to catch up with the US/China+ models. This would provide a significant incentive for countries within US/China+ to engage with the Path 2 negotiations early on with the aim of reaching an agreement that impacts all signatory countries, whilst their own AI models/companies are still ahead. Just as it is an incentive for the US/China+ to finalise the global AI governance negotiations, there is an incentive for other countries, especially those with the leading AI capability/models to delay finalisation of negotiations. It is likely, however, that the US and China will be powerful players in the supply chains of competing Non-Path 1

²⁰ <https://arxiv.org/abs/2308.15514v2>

²¹ <https://arxiv.org/abs/2511.10783>

²² <https://controlai.org/uk-asi-bill>

companies, and as such, they would be well placed to exercise control over any other country that was abusing the situation.

The aim would be for the US and China to join Path 2 negotiations as soon as they both (not just one) agree to do so, along with any other states in US/China+ not already participating in Path 2.

F. A Pro-Human vision of AI

As society starts to engage more and more with Artificial Intelligence, a strong counter swell of opinion has developed, particularly in the United States. In September 2026 the Future of Life Institute brought to Washington leading representatives of that ground swell including Steve Bannon from the far right and Bernie Sanders from the far left, speaking on the same platform. They were advocating a Pro-Human vision for AI as an alternative to the Race to Replace vision being pursued by Silicon Valley. The difference between these two visions is profound, both in terms of Frontier AI Safety, but also the wider impacts of AI upon society.

The struggle between Pro-Human supporters and Race to Replace supporters will initially be fought out in Path 1, then actively engage with Path 2 as it gets underway.

G. Overall process

AI is an increasingly powerful technology which is already having a transformational impact on the world. The technology is advancing far, far faster than the governance and regulation to control it, and this imbalance must be rectified.

The remedy proposed is to proceed down two distinct paths, one based upon safety and security as a matter of particular urgency and the other based upon developing the other elements of a global AI agreement ahead of the final negotiation of a truly global AI agreement.

The steps would be

1. Path 1 Frontier AI Safety / Security
 - a. Development of tools and procedures that could be used to pace frontier AI is prioritised
 - b. Governments with frontier AI models in their countries need to ensure the effective evaluation and release control of the models
 - c. They may decide to pace the development of the frontier models in their own countries.
 - d. In parallel, the US and China should seek to work together (and, if they so choose, with other countries with frontier AI companies) to control frontier AI (US/China+)
 - e. US/China+ should agree on the best approach to pace frontier AI
 - f. US/China+ should agree to implement that approach
2. Path 2 Global AI Governance
 - a. Meanwhile, a coalition of Middle Powers needs to be formed to prepare a much broader agreement on AI governance
 - b. They will need to establish the scope of what needs to be agreed at the global level and what can be successfully managed at the national level
 - c. They can start to explore those issues that need to be agreed at the global level, such as an Agency and a licencing system
3. Single united path:
 - a. When both the US and China wish to participate in the negotiation of a truly global AI agreement, then a full negotiation can commence
 - b. The negotiations would start from scratch, but the output of the Path 2 work would be available to enable the process to move much faster than would otherwise be the case

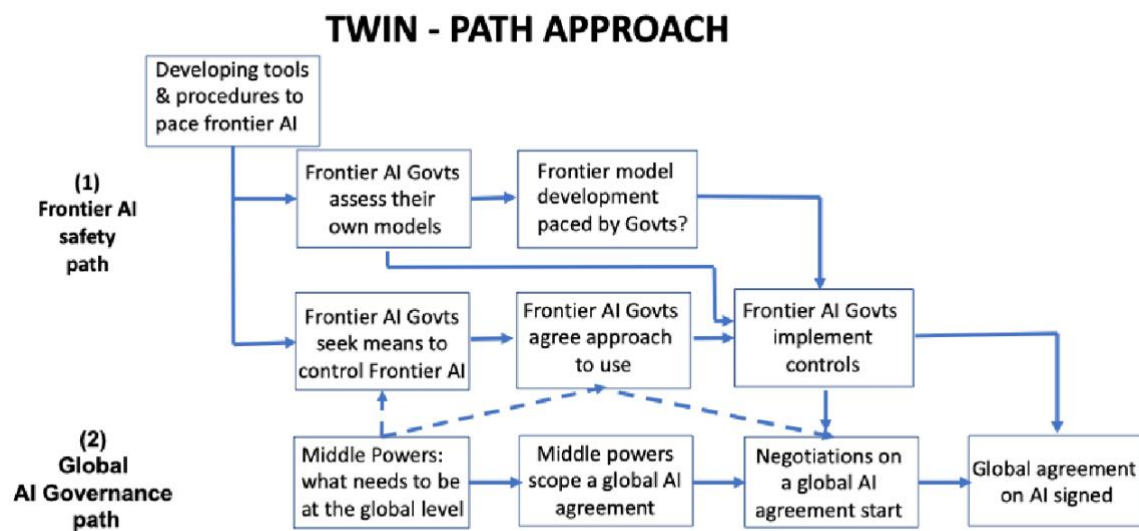


Figure 1

H. Conclusion

To conclude, Path 1 buys time, whilst Path 2 uses the time that has been bought to take preparatory steps towards a truly global agreement on AI governance (see Figure 2).

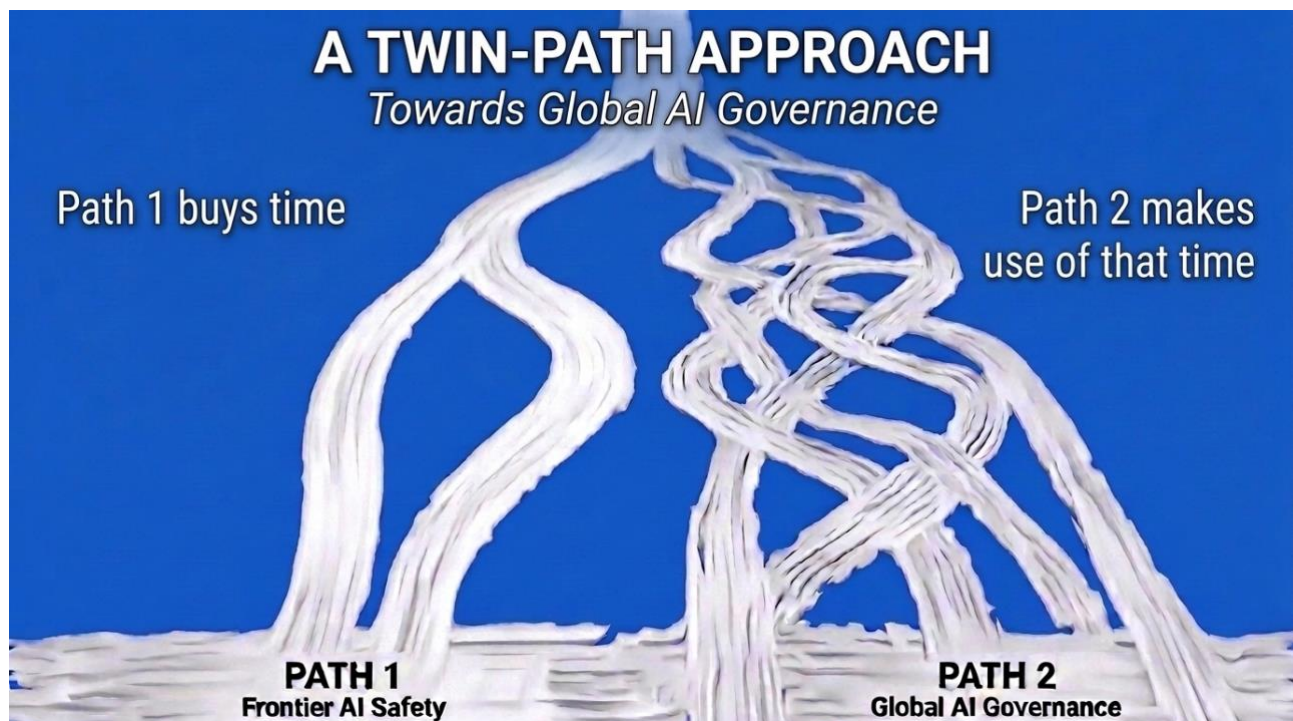


Figure 2

This is a report by the Global AI Governance Alliance.

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 Robert Whitfield

Convener of the Global AI Governance Alliance
London, United Kingdom

Suggested Citation

Whitfield, R., 2026 Twin-Path Approach to AI Global Governance
Global AI Governance Alliance

A TWIN-PATH APPROACH TO AI GLOBAL GOVERNANCE

Global AI Governance Alliance | www.gaiganow.org