



THE  
FUTURE  
SOCIETY

# 2022 ANNUAL REPORT

---

# CONTENTS

---

<b>President's letter</b>	3
<b>Our independence</b>	5
Board	5
Finances	6
<b>Our 2022 Achievements</b>	7
<b>AI and Rule of Law</b>	7
Convening the Fourth Edition of The Athens Roundtable on AI and the Rule of Law	7
Launch of Global Online Course on AI and the Rule of Law	9
Publishing A Manifesto on Enforcing Law in the Age of AI	10
Strengthening Coordination for Interoperable Benchmarks	11
Producing a Model Protocol for Assessment and Validation of AI in Legal Discovery	11
<b>European AI Governance</b>	12
Advocating for the Governance of General-Purpose AI Systems in the EU AI Act	12
Championing Regulatory Sandboxes in the EU AI Act	14
Enforcement of the EU AI Act Reflection Group	14
<b>Global &amp; Corporate AI Governance</b>	15
Developing National AI Strategies and Implementation Plans for Tunisia and Ghana	15
Co-producing Global Reports on AI and the Environment	17
Supporting the Operationalization of Global Normative Frameworks on AI	18
<b>Our 2023 Outlook</b>	20
Our guiding tenets	21
Looking ahead	22
Why donate to TFS?	24
<b>Contact us</b>	27



# PRESIDENT'S LETTER

## ***The race toward general-purpose AI systems is on, and our institutions are not ready***

Over the past fifteen months, we have all been shocked and awed by the speed and scale at which new general-purpose AI systems have been developed and deployed, demonstrating astonishing leaps in capabilities. It is as if we surged past the “Turing test”, without even realizing it! As they are scaled up from billions to trillions of parameters, these large pre-trained transformer models span an increasing set of modalities and cognitive functions, from language to vision and now reasoning; they are beginning to be able to perform more and more complex and integrated tasks in digital and physical environments.

As I pen this letter, it has been just over two weeks since GPT-4 has been released. Researchers have merely begun to gauge and discover the model’s potential and emerging capabilities. Only a few years ago, many of the AI systems that exist today seemed to be decades away, if not mere figments of imagination. We are approaching the brink of “artificial general intelligence” much quicker than we all thought. Severe economic, social, and political turbulence looms ahead.

Our institutions are not ready. Nor do we have in sight the global governance regime needed to build requisite safeguards and address critical uncertainties regarding the access, control, and ownership of these technologies. The growing tensions in Europe—with the unprovoked and illegal invasion of Ukraine—across the Gulf, as well as in East Asia, do not bode well for international cooperation on these critical matters, not to mention the mounting climate emergency which adds techno-solutionist pressure.

Unprecedented sums of capital are now being invested in supercharged General-Purpose AI Systems (GPAIS) labs as core digital markets—such as search engines—seem to be ripe for disruption. Whereas the market for conversational agents was in shambles just less than 6 months ago, ChatGPT has unveiled countless new market opportunities and productivity which innovators are racing to materialize. But as the race between big tech platforms and their partner AGI labs accelerates to a point of recklessness, the underlying techno-scientific paradigm imposed on society remains risky: these powerful general-purpose AI systems remain uninterpretable black boxes; we do not yet have a clear, scientific pathway to resolve the challenges of interpretability, control, and alignment.

Hence the urgent need to shape market incentives and behaviors in a positive direction, through commonsense governance of the research, development, and deployment of GPAIS. Focusing only on their application and use will not be sufficient. Given that this technology is “general-purpose” in nature, and that most of its value and related risks are generated at the “pre-training” stage, we need to have the courage to climb upstream in the value chain to hone in on the risks posed in their research and development.

Will we be able to tame the powerful beasts we are developing and prevent large-scale accidents, misuse, social unrest, or malicious use while allowing innovation to flourish? I am deeply concerned that we still lack adequate institutional instruments, mechanisms, and related capacities and practices. This represents an immense collective action challenge. Given the deep-seated misunderstandings and growing polarization across the AI governance community, forging common ground toward resolute and sustained action will not be easy. It will require humility, dialogue, and willingness to compromise on all sides. We at TFS are committed to it.

I am glad to share that, in this regard, this past year was also a pivotal time for The Future Society (TFS), as we close the year better positioned to achieve our mission to align AI through better governance. Following last year's strategic reboot, we implemented a new operational model, allowing us to work in a more agile manner, and to allocate our resources more strategically towards the AI governance projects that have the greatest potential for positive impact. Moving forward, we will focus directly on the governance of GPAIS (or "foundation models," in some contexts), particularly on their research and development, a neglected area in AI governance and one where our unique skillset can help ensure these systems are safe and aligned with fundamental human values.

This year's annual report demonstrates how this strategic shift is coming to fruition. Following our policy research and advocacy efforts to support the 'future-proofing' of the EU AI Act, a number of our institutional innovation recommendations have been considered in legislative drafts of the Act. As ensuring that the global south is part of the conversation is key to ensuring that any governance regime is legitimate and viable over time, our team collaborated with government ministries and a diverse range of stakeholders in Ghana, Rwanda, and Tunisia to develop National AI strategies that are now being implemented by their respective governments. We ended 2022 with the fourth edition of The Athens Roundtable, which was held this year at the European Parliament in Brussels, and supplemented by polls, questions, and discussions amongst an online audience. This year's edition reached an audience of policymakers, legal experts, judicial operators, civil servants, industry leaders, and civil society representatives from 112 countries. In keeping with technological and regulatory trends, discussions spanned the forthcoming EU AI Act, the general-purpose and generative AI revolution, and the coordination of international efforts to measure and evaluate AI systems, among many other timely topics.

Having played a role in building the field of AI governance for the past decade, it has been awe-inspiring to see AI governance, once a niche interest, permeating policy discourse and percolating into mainstream media. I would like to express my deepest gratitude to those who have supported the work of The Future Society, be it financially, in-kind, or otherwise, through the years of growth of our organization and the maturation of our field.

Finally, we are eager to scale our impact. In the coming year, we have a number of prospective projects we wish to launch which require growing our team. You may read this report's section on Funding Gaps to learn more about how we intend to extend our impact, and, if this resonates with you, I humbly request that you consider becoming a supporter of TFS.

Sincerely,

*Nicolas Miaillhe*

President and Founder  
The Future Society (TFS)





---

## OUR INDEPENDENCE

---

### Board of Directors

Directors have distinguished backgrounds in academia, government, business, or civil society, and are responsible for decisions that affect the strategy, finances, and operations of the organization. All Directors operate in a volunteer capacity.



**De Kai**  
Professor of CS  
and Engineering,  
HKUST



**Julien Merceron**  
Worldwide CTO,  
Bandai Namco  
Studios



**Nicolas Mialhe**  
Founder &  
President, The  
Future Society



**Sarah Pearce**  
Partner, Privacy and  
Cyber Security  
Practice, Paul  
Hastings

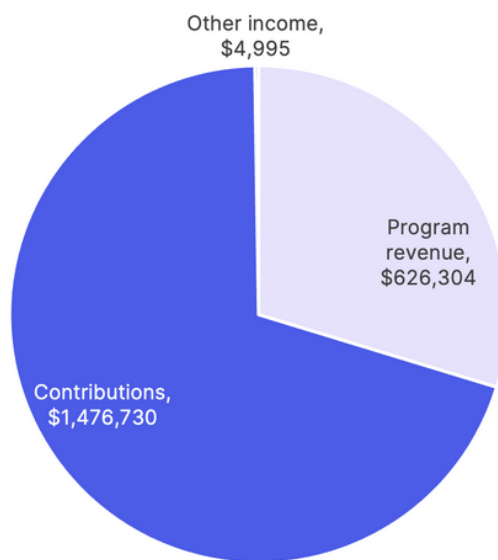


**Simon Mueller**  
Co-Founder &  
CEO, Mantaro  
Partners

## Finances\*

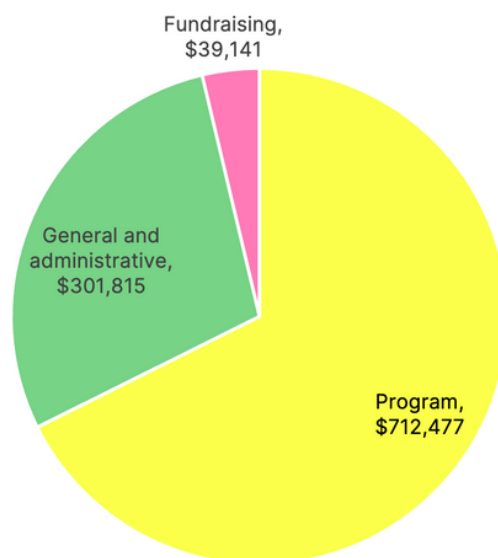
The Future Society is an independent organization primarily supported by contributions from philanthropic organizations, and funded in part through donations and service-based contracts with IGOs, governments, and private organizations. Contributors are not provided “substantial return benefit” as defined in U.S. Internal Revenue Code § 513 and accompanying regulations. More details can be found on Form 990s accessible on our site.

### Revenues



Total: \$2,108,029

### Expenditures



Total: \$1,053,433

\*Figures are pre-audited. Totals will be updated, as necessary, upon completion.



## OUR 2022 ACHIEVEMENTS

### AI AND RULE OF LAW

We aim to ensure that the research, development, and deployment of advanced AI is conducted in a manner that protects and upholds the rule of law.



### Convening the Fourth Edition of The Athens Roundtable on AI and the Rule of Law

In early December 2022, the fourth edition of The Athens Roundtable on Artificial Intelligence and the Rule of Law brought together policymakers, legal experts, judicial operators, civil servants, industry leaders, and civil society representatives from 112 countries. Hosted physically at the European Parliament in Brussels, the Roundtable welcomed 120 onsite and over 1,000 virtual participants.

This year's Roundtable was held under the auspices of Her Excellency, the President of the Hellenic Republic Ms. Alkaterini Sakellariopoulou, and was organized in partnership with the Patrick J. McGovern Foundation. The event was proudly co-hosted by prominent intergovernmental organizations, including the European Parliament, UNESCO, OECD, and the Council of Europe, as well as leading AI institutions and firms, including IEEE, Cravath, and Amazon Web Services. It was generously supported by the Jain Family Institute, Arnold & Porter, and Debevoise & Plimpton.



Co-founded by TFS in 2019, The Athens Roundtable has served as the premier international, multi-stakeholder AI governance forum within the prism of legal, judicial, and compliance systems. The fourth edition focused on the implementation and enforcement of laws, regulations, standards, and policies across the industrial value chain of AI systems. In keeping with technological and regulatory trends, topics spanned the forthcoming EU AI Act, governing generative AI, and coordinating international efforts to measure and evaluate AI systems.

[Watch the recordings >](#)
[Read the report >](#)

*"The Future Society has been UNESCO's valuable partner in the realization of the Recommendation on the Ethics of Artificial Intelligence, to advance AI ethics and governance globally. Through The Athens Roundtable and participation in international expert groups, TSF is adding a powerful voice to build regulatory frameworks, reinforce institutional capacities, and advocate for diversity and inclusion at the core of decision-making processes, which is also core to UNESCO's efforts. TFS is actively shaping the implementation of governance mechanisms, standards-setting, evaluation and measurement for trustworthy AI. Their pioneer work with strategic foresight on general-purpose AI systems and regulatory sandboxes is a great example that can positively influence the safeguard of fundamental rights in the age of AI. At UNESCO, we are proud to partner with them."*



**Gabriela Ramos**

Assistant Director-General for the Social and Human Sciences of  
UNESCO



## Launch of Global Online Course on AI and the Rule of Law

In March 2022, The Future Society and UNESCO—with the support of the National Judicial College, CETIC.br/NIC.br, and IEEE SA—launched a global Massive Online Open Course (MOOC) on AI and the Rule of Law. Structured around six modules, the course provides an introduction to the opportunities and risks of the adoption of AI across judicial systems and explores AI's impact on the administration of justice.

The pilot course welcomed 4,432 participants from over 140 countries, delivering on its objectives to strengthen the capacities of judicial operators worldwide—to protect human rights, democratic principles, and the rule of law.



The course was developed by The Future Society and UNESCO, with the support of Cetic.br|NIC.br, the National Judicial College, and IEEE SA. It was made possible by the kind contributions of the UNESCO Multi Donor Programme on Freedom of Expression and Safety of Journalists (MDP) and the Open Society Foundations.

[Read more >](#)

[Enroll in the course >](#)



*"Of the many lessons learned, I would point out as the most relevant the importance of 'molding' the use of AI within the judiciary to our core ethical values through a human rights-based approach. If we as judges choose to 'turn our face' and let the 'tech guys' handle it, ignoring the way AI will be implemented within the judiciary, our future generations will have us to blame for the probable distancing the decision-making of tomorrow will have from our humanistic principles we appreciate today."*

**Sergio Torres Teixeira**

Justice in the 6th Regional Labour Court of the Brazilian Labour Justice

## Publishing A Manifesto on Enforcing Law in the Age of AI

From June to October 2022, a working group under the title of the Transatlantic Reflection Group on Democracy and the Rule of Law in the Age of “Artificial Intelligence,” comprised of legal practitioners and academics from both sides of the Atlantic, convened to draft A Manifesto on Enforcing Law in the Age of “Artificial Intelligence”. This Manifesto calls for the effective and legitimate enforcement of laws concerning AI systems, and, in doing so, recognizes the important and complementary role of standards and compliance practices.



The Manifesto offers 10 recommendations to address enforcement challenges shared across transatlantic stakeholders. These recommendations present practically feasible steps that policymakers and regulators can take to create enforcement regimes that are both effective throughout the lifecycle of AI systems and consistent with democratic value



The Manifesto was subsequently discussed at a fireside chat at The Athens Roundtable in December 2022 with US FTC Commissioner Rebecca Kelly Slaughter and Principal Advisor to the European Commission’s Directorate-General for Justice and Consumers Paul Nemitz, moderated by Dr. Francesco Lapenta.

[Watch the fireside chat >](#)

[Read and sign the Manifesto >](#)





## Strengthening Coordination for Interoperable Benchmarks

TFS launched an interoperable benchmarking working group under The Athens Roundtable that brings together members from major standards organizations and technical communities, including U.S. NIST, CEN-CENELEC, IEEE, LNE, VDE, and Greece's National Centre of Scientific Research. This transatlantic working group initiated "A Proposal for Advancing the Responsible Use of AI through Benchmarks," a paper integrating a range of responsible AI development and deployment guidelines to propose an institutionalized, interoperable, open, and ongoing AI benchmarking program (expected to be published in 2023). Through this working group, TFS has also explored the possibility of a joint transatlantic roadmap on AI measurement/metrology, echoed in the US-EU Trade & Technology Council's Joint Roadmap on Evaluation and Measurement Tools for Trustworthy AI and Risk Management, which called for advancing shared terminologies and taxonomies.



## Producing a Model Protocol for the Assessment and Validation of AI Systems Used for Legal Discovery

Throughout 2022, in a workstream led by Dr. Bruce Hedin, TFS produced a model protocol for the assessment and validation of AI systems used for Technology-Assisted Review (TAR) in legal discovery. The protocol is grounded in statistical methods and is interoperable—designed in such a manner that the techniques can be used to assess AI systems used in domains outside of legal discovery. Preprints of the Protocol, Handbook, and Commentary have thus far received favorable reviews from influential legal experts, representatives of the National Judicial College, the U.S. Federal Judicial Center, NYU, and a Former US magistrate judge. These documents are expected to be published by TFS and IEEE in 2023.

## EUROPEAN AI GOVERNANCE

We aim to strengthen the proposed EU AI Act and the surrounding regulatory regime, at large, instilling precaution with regard to increasingly capable AI systems.



### Advocating for the Governance of General-Purpose AI Systems in the EU AI Act

Over the past few years, we have witnessed the emergence of “general-purpose AI systems” (GPAIS)—AI systems that are able to perform generally applicable functions and able to have multiple intended and unintended purposes. We believe that GPAIS pose risks that are qualitatively different and more severe than those of more conventional, narrow AI systems.





In an effort to ensure that the EU AI Act is robust to future advancements in AI capabilities, The Future Society has promoted and advocated for a special governance regime tailored to GPAIS. This includes institutionalized dialogue between frontier labs and independent authorities, the development of benchmarking capabilities to assess generality and safety/alignment, and an efficient and effective distribution of responsibilities across the GPAIS value chain.

TFS has provided policy advice on how to best govern AI, AI supply chains, and GPAIS. We have produced 9 tailor-made policy briefings and have conducted more than 60 meetings with over 25 policymakers and advisors at the heart of AI Act negotiations.

*"Over the years, TFS has shaped the development of AI governance for the benefit of humanity through their careful research and forward-thinking, evidence-based advocacy across the Atlantic. By advising on the EU AI Act, convening U.S. and EU decision-makers, and advocating for general-purpose AI systems regulation, the team has made an impact on the development of a safer and more ethical AI governance."*

**Anthony Aguirre**

Vice President & Secretary of the Board  
Future of Life Institute





## Championing Regulatory Sandboxes in the EU AI Act

TFS has been advocating for regulatory sandboxes to be implemented in the EU legislative ecosystem, beginning in early 2020 with the publication of *Experimentation, testing & audit as a cornerstone for trust and excellence* in response to the European Commission's February 2020 White Paper on Artificial Intelligence.

In February 2022, TFS published a memo titled *Sandboxes without the quicksand: making EU AI sandboxing work for regulators, entrepreneurs and society*. This memo proposed a three-stage program aimed to help policymakers implement regulatory sandboxes. Following that, in July 2022, regulatory sandboxes were institutionalized in Europe as part of the European Commission's New European Innovation Agenda.

We welcome the European Parliament's decision to carry our recommendations regarding the design and operationalization of sandboxes into negotiations. We look forward to continuing to contribute to the development of regulatory sandboxes, among other innovative elements of the EU AI Act.



[Read our policy memo >](#)

## Enforcement of the EU AI Act Reflection Group

The mechanisms by which the AI Act it will be enforced remain up for deliberation. Some are concerned that its compliance may be reduced to a “box-ticking” exercise—as we have witnessed to some extent with the GDPR—or to rely heavily on ex-post adjudication.

This Reflection Group, which TFS initiated in 2022, aims to contribute to the development of enforcement mechanisms by pooling the collective intelligence of involved decision-makers from the Council and the Parliament. We work closely with members to understand their preferences and priorities, identify the cruxes of disagreements, and conduct research to support informed dialogue. In this manner, we arrive at more feasible and effective enforcement approaches. This work is projected to continue throughout 2023.

## GLOBAL & CORPORATE AI GOVERNANCE

We support the development and implementation of agile governance frameworks for AI, ranging from international frameworks to corporate guidance and technical standards.

### Developing National AI Strategies and Implementation Plans for Tunisia and Ghana

In 2022, TFS supported the development of Tunisia's and Ghana's National AI Strategies in collaboration with public sector ministries, GIZ, and Smart Africa. TFS led participative consultations with local institutions, conducted SWOT analyses of local AI ecosystems, drafted ethical guidelines, and developed implementation plans to support each country in achieving its respective national objectives and sustainable development goals.



In June 2022, TFS delivered The National Artificial Intelligence Strategy of Tunisia with our partners, Tunisia's Ministry of Industry, Mines and Energy, Ministry of Communication Technologies, Ministry of Higher Education and Scientific Research, Ministry of Economy and Planning, and GIZ Digital Transformation Center Tunisia (BMZ). Consultations in Tunis ultimately informed the design of policy pillars, ranging from human capital development to sectoral adoption, cloud computing strategies, and data governance.





In October 2022, TFS delivered the Republic of Ghana National Artificial Intelligence Strategy 2023-2033 with our partners, Ghana's Ministry of Communications and Digitalisation and Data Protection Commission, GIZ FAIR Forward, and Smart Africa. The strategy was informed by over 40 local stakeholder consultations, AI policy landscape mapping, and an in-depth analysis of the strengths, weaknesses, opportunities, and threats (SWOT) of Ghana's AI ecosystem. Four high-level public sector consultation workshops, including a transnational peer-learning workshop with experts from Rwanda—with whom we collaborated to develop Rwanda's national AI strategy—were held to refine the action plan and to produce a detailed booklet of use cases across critical sectors. The resulting cross-ministerial and cross-sectoral policies aim to promote responsible, sustainable, and AI adoption in Ghana.



*"[GIZ] FAIR Forward in Ghana have had the pleasure of working with The Future Society on The Ghana National AI Strategy and have been extremely impressed with their knowledge, skills, and professionalism. TFS has a deep understanding of the legal, ethical, and policy issues related to AI and has provided invaluable guidance and support to us. Their team of experts has helped us navigate the complex landscape of AI policy and governance, and has provided us with strategic recommendations that have enabled us to make informed decisions. In addition to their technical expertise, TFS has also demonstrated exceptional communication and interpersonal skills. They have been responsive to our needs and have consistently delivered high-quality work on time and within budget. Their team has also been proactive in anticipating our needs and providing us with timely updates on the latest developments in AI policy. We very much appreciate the level of professionalism, dedication and excellence TFS demonstrated during this process and look forward to future collaborations and partnerships."*

**Emmanuel Mumuni**

Programme Component Manager, GIZ Ghana





## Co-producing Global Reports on AI and the Environment

In his capacity as Co-Lead of the Global Partnership on AI's Project RAISE (Responsible AI for the Environment), TFS President Nico Mialhe led the commissioning of three actionable reports in 2022. These reports were presented at the Sharm el-Sheikh Climate Change Conference (COP 27) in November, and the United Nations Biodiversity Conference (COP15) in Montreal in December.

- A pioneering report on the use of responsible AI for biodiversity conservation, Biodiversity and Artificial Intelligence: Opportunities and recommendations for action.
- A joint scoping report with OECD, Measuring the environmental impacts of Artificial Intelligence compute and applications: The AI footprint, which outlines a path toward benchmarks, frameworks, and impact assessments for the direct and indirect environmental impacts of AI compute.
- A booklet designed to help companies in the power sector understand the prerequisites for deploying AI systems to facilitate a transition to net zero electricity consumption, titled AI for Net Zero Electricity.

[Read the report >](#)[Read the report >](#)[Read the booklet >](#)

## Supporting the Operationalization of Global Normative Frameworks on AI

Throughout 2022, TFS supported the development and implementation of global coordination mechanisms and normative frameworks on AI.



### **Organisation for Economic Co-operation and Development (OECD)**

As a key contributor to the development of the OECD AI Principles throughout 2018-2019, TFS continues to support the work via the AI Governance Working Party (AIGO). Launched in 2022, AIGO brings together 38 countries (and many observers) to coordinate their AI policy and regulatory approaches. In 2022, TFS was involved in projects pertaining to the classification of AI systems, the long-term impact/risks of AI systems, coordinating national AI policy frameworks, compute measurement and monitoring frameworks, and regulatory sandboxes. These projects play an important role in building the global influence of emerging legislative and regulatory instruments, such as the EU AI Act.



### **United Nations Educational, Scientific and Cultural Organization (UNESCO)**

TFS staff have been appointed to multiple advisory expert groups on AI ethics and international AI policy. In 2019-2021, TFS was involved in the development of the Recommendation on AI ethics, adopted by 193 countries in November 2021. In 2022, TFS team members were invited to contribute to the High-Level Expert Group on Implementing the Recommendation on AI Ethics, which is currently developing a toolkit comprising a Readiness Assessment Methodology (RAM) and an Ethics Impact Assessment (EIB).



### Global Partnership on AI (GPAI)

GPAI, which was initially launched by the G7 countries to lead the global governance of AI, now includes 29 countries. In 2020 and 2021, TFS won three competitive open tenders and delivered advisory reports on responsible AI and AI for COVID-19 pandemic response to scope GPAI's agenda on these topics. In 2022, President Nico Mialhe co-led the Committee on Climate Action and Biodiversity Preservation and served on the initiative's Steering Committee.



### Standard-setting Organizations (SSOs)

In 2022, Several TFS staff served as members of technical standardization committees, notably SC 42 Artificial Intelligence Committee at ISO, and its mirror committee at CEN-CENELEC—which shapes industry standards related to AI risk management, verification, and validation.



### U.S.-EU Trade and Technology Council (TTC)

One TFS staff member is a member of the Council and other members of our staff been invited to join the US-EU discussions on standardization and benchmarking of AI and governance of AI supply chains.





## OUR 2023 OUTLOOK

In 2022, we witnessed rapid advancement in AI capabilities, including breakthroughs in protein sequence prediction, forays into multimodal task completion, and instances of AI systems using cooperation and deception in natural language to beat humans in complex games.

Such emergent capabilities are made possible by technological advancements in “general-purpose AI systems” (GPAIS)—large, machine-learning models (in some contexts referred to as “foundation models”) trained on vast quantities of data and that can be adapted to a wide range of tasks. Due to their versatility, and the advantages they confer in terms of business and productivity, GPAIS have captured not only the attention of technical research communities but have also become a frontier for AI startups racing to develop “artificial general intelligence.”

We believe that the emergence of GPAIS may foreshadow the most disruptive technological revolution in human history. At TFS, we are concerned that these advanced AI systems models pose serious risks to human safety and fundamental values. We also believe that relative to the threat that they pose—and in part due to their novelty—GPAIS have been neglected in the AI governance landscape. Thus, in the coming year, we intend to orient our work to focus on the governance of GPAIS, and in particular, towards the research and development phases of such models—stages of the AI lifecycle where risk mitigation is most tractable, and preventative measures would be most effective.

## Our guiding tenets

### Inclusion

We believe that legitimate and sustainable governance requires bringing to the table many different perspectives. When teams lack diversity—in academic, social, ethnic, or political backgrounds—their outcomes tend to overlook critical perspectives and perpetuate socioeconomic inequalities. Ours is a field of considerable homogeneity. We make an effort to counterbalance this by engaging with underrepresented communities, working with teams and individuals representing a broad range of stakeholders—including governments, intergovernmental organizations, the private sector, academia and civil society, and by creating a professional environment in which all individuals' identities and preferences are respected.

### International Coordination

We believe that effective governance requires international dialogue and coordination among nations. The AI revolution is a global phenomenon. Though there are development "hot spots," the impacts of these technologies are not constrained by geographic boundaries or political borders. To effectively address the large-scale risks posed by advanced AI systems, we believe we will require an unprecedented degree of cooperation and coordination between governments and intergovernmental institutions.

### Future-proofing

We orient our work to address both near- and long-term implications of AI. We aim to not only address the current needs and challenges posed by AI, but also to anticipate and prepare for its future, far-reaching impact on various aspects of society. By taking a forward-thinking approach, we strive to create solutions that will be relevant and sustainable in the long run.

### Practicality

We prioritize ideas by their potential to achieve real-world impact. AI research and development takes place at a lightning-fast pace; there can be large shifts of capital and technological advantage overnight. Due to these power dynamics, we believe that certain governance "nodes" may be significantly more impactful or tractable than others, but we acknowledge that—again, owing to the dynamics of this field—we face many uncertainties when trying to determine which governance approaches are robustly lead to positive outcomes. This is why we have developed a methodology to steer the allocation of our resources towards approaches we have determined are the most likely to have a positive impact.

### Innovation

We think outside the box to develop solutions for operationalizing governance, putting institutional innovation at the center of our methodology. We are concerned that governments are not sufficiently equipped to mitigate the risks posed by advanced AI systems. In a field characterized by rapid capital influx and technological diffusion, many governments lack the means to measure and monitor the impact that AI systems will have on individual, sociological, and economic levels. Standing to this task will require innovation and adaptation at the institutional level—in some cases, institutions may need to adapt to new functions, or entirely new institutions may need to be formed, to support the structural changes needed for effective AI governance across sectors and geographies.





## Looking ahead

In 2023, we plan to continue a number of existing projects and launch new projects that develop, advocate for, and facilitate the implementation of voluntary and non-voluntary governance mechanisms that address the risks posed by advanced AI systems. As explained above, we intend to orient our focus to address risks that may emerge in the research and development of general-purpose AI systems (or in some contexts, "foundation models"). The following is a preview of our 2023 portfolio of activities.

### AI AND RULE OF LAW

It will be essential that the rule of law remains respected and upheld throughout turbulent times of technological diffusion. Beginning in 2023, we intend to strengthen our policy engagement efforts in the U.S., in an effort to build the capacity of the U.S. Government, on both federal and state levels, to effectively govern advanced AI systems. In doing so, we will continue to promote coordination between regulators and policymakers on both sides of the Atlantic.

[More about this theme >](#)



## EUROPEAN AI GOVERNANCE

In 2023, we will build upon our 2022 activities to improve the development and enforcement of the EU AI Act by advocating for precautionary risk mitigation measures that address general-purpose AI systems (GPAIS). Our work plan includes continued advocacy for the implementation of regulatory sandboxes, an in-depth analysis of the AI Act's governance rules for GPAIS, and the development of enforcement mechanisms through an established reflection group that involves key European policymakers.

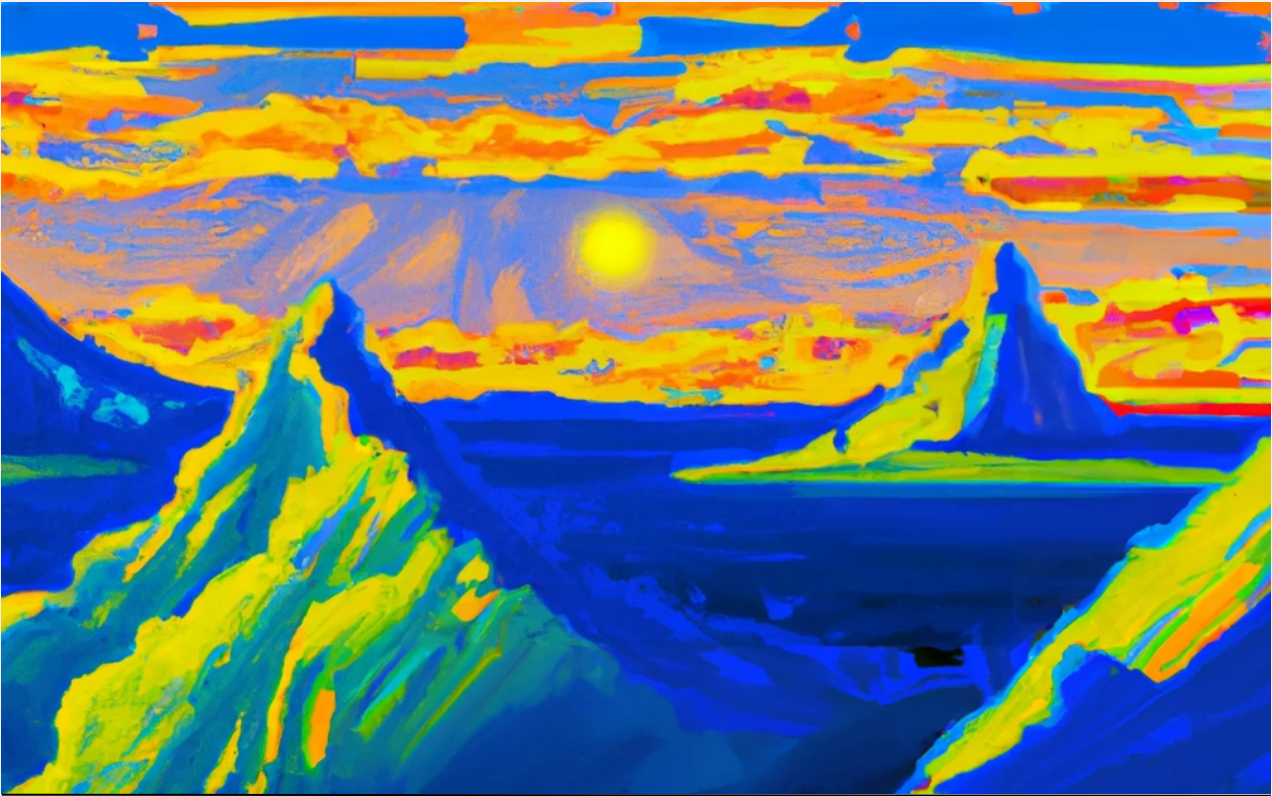
[More about this theme >](#)



## GLOBAL & CORPORATE AI GOVERNANCE

As with many of today's major governance challenges, from climate change to pandemic response, AI governance requires multi-stakeholder, transnational, and cross-sectoral coordination. This thematic area includes the development of corporate governance frameworks, contribution to safety-enhancing benchmarking and metrological methods, and advisory to leading international organizations (e.g. OECD, UNESCO, GPAI) and standards-setting bodies whose remits concern the governance of AI. Beginning in 2023, we are orienting towards focusing on risks present in the research and development of advanced AI systems, as we believe that this is the stage of the AI lifecycle that is most tractable in terms of reducing the risk of large-scale incidents.

[More about this theme >](#)



## Why donate to TFS?

**1. AI governance is important.** Artificial intelligence is transforming the human experience. Some experts believe that AI will radically improve the world; some believe it will open doors to catastrophic incidents. Many experts, in fact, believe both scenarios are possible. One thing is certain: our future is will be the result of human decisions to design, test, and deploy technology. This, we believe, is the collective action problem of our time!

**2. AI governance is tractable.** Humans are behind the wheel, and changes in human behaviors can steer the course of technological development toward a safer and more equitable future. Responsible investments in AI metrology and safety-oriented research, for example, can help us prepare better-informed regulatory frameworks before more powerful systems are deployed. Public-private cooperation can provide the resources and perspectives needed to tackle the complex task of overseeing increasingly capable AI systems. Industry-wide responsible practices and legislative action can slow the unfettered deployment of unsafe systems. Governments can establish verifiable, enforceable agreements to prevent large-scale hazards. It will require a concerted effort of researchers, policymakers, and companies to ensure that AI systems of the future align with human values and the rule of law.



**3. AI governance may be a front-loaded challenge.** Laws, policies, and norms have compounding effects. While crafting AI governance will be an ongoing process, getting governance mechanisms in place early is critical. This will necessitate investing heavily in the ‘upfront’ work—conducting research, developing policies and regulations, establishing standards and best practices, and engaging with stakeholders—needed to lay foundations for more robust laws, policies, and norms in the long term, even when many difficult questions remain unanswered. Although this front-loaded effort is onerous, failing to prepare for the technology of the future could prove far more costly.

**4. Our work can have a long-term, positive impact.** The institutions, laws, and norms we establish now will shape the development and use of AI systems for years to come. A proactive effort could steer technology towards benefiting society broadly. On the other hand, if we fail to prepare adequately or act with foresight, we risk stumbling into a future in which powerful AI systems cause irreversible damage to our social fabric. How we research, discuss, and craft governance frameworks in this early stage may reverberate for decades. This makes the work of AI governance impactful, even if it can feel like just the first steps.

**5. We take a prioritized, portfolio-based approach.** Our resources are limited, but the challenges and opportunities in AI governance are vast. This is why we prioritize a ‘portfolio’ of work that tackles the most significant and tractable problems. By methodologically assessing the landscape of governance approaches, we aim to focus our efforts on the areas of the greatest potential impact. For example, we may address understudied gaps or build on emerging best practices with multiplier effects. A portfolio approach also allows us to balance diverse work streams, such as researching and advocating for safer industry practices, convening policymakers, companies, and civil society to hone in on fundamental governance challenges, and providing capacity-building and educational programs for judicial operators and the public at large.

**6. We are independent.** As an independent organization not beholden to companies or governments, we are uniquely well-positioned to work toward AI governance that serves the public interest. We aim to bring impartial analysis and judgment to governance discussions without profit motives or political pressures distorting our work. Our non-affiliated status enables us to function as a neutral convener and collaborator with external stakeholders, including companies, policymakers, and other actors working towards responsible stewardship of AI.

**7. We face existing funding gaps.** We have a number of potentially high-impact activities queued up, but due to funding constraints, we are currently unable to launch them. By funding our work, you could help us make progress on a range of open questions—and ultimately, help determine whether advanced AI is shaped by commercial interests alone or by a diversity of voices.

**8. We are fighting an uphill battle.** Corporations are pouring billions of dollars into AI, and their profit incentives can be misaligned with broader public interests. Independent nonprofit organizations are crucial players to ensure that governance solutions serve the public interest, and we rely upon philanthropic support to have an effect commensurate with other stakeholders' influence.

**9. We have a strong track record.** Since our incorporation in 2016, TFS has been a major player in the production of foundational AI governance frameworks and policies. Here is a short list of our accomplishments:

- Leadership in developing the [OECD AI Principles](#) and [Framework for the Classification of AI Systems](#), among other leading multilateral frameworks for the global governance of AI
- Convened over 5,000 regulators, policymakers, judicial operators, and AI developers through four editions of [The Athens Roundtable on AI and the Rule of Law](#) and the [Global Governance of AI Roundtable](#) at the World Government Summit
- Trained over 4,500 judicial operators through our [Massive Online Open Course \(MOOC\) on AI and the Rule of Law](#)
- Led advocacy for [regulatory sandboxes](#) to be introduced to the EU AI Act—adopted into European Parliament's Committee on Industry, Research and Energy draft opinion on the EU AI Act.
- Completed [3 projects](#) and helped commissioned 3 projects with the Global Partnership on AI, and led the development of [3 national AI strategies](#) in the Global South

[Make a donation to TFS >](#)



*"The Future Society has demonstrated an inspiring commitment to shaping and scaling governance solutions for AI alignment. The team stands out for their convening power—bringing together relevant stakeholders from industry, academia, and the public sector and steering pivotal discussions on AI governance challenges at The Athens Roundtable on AI and the Rule of Law."*

**Connor Leahy**  
CEO, Conjecture





# THE FUTURE SOCIETY

---

## CONTACT US

---

General: [info@thefuturesociety.org](mailto:info@thefuturesociety.org)

Press: [press@thefuturesociety.org](mailto:press@thefuturesociety.org)



The Future Society  
867 Boylston Street, 5th Floor, Boston MA 02116, United States  
[www.thefuturesociety.org](http://www.thefuturesociety.org)