AI ACTION SUMMIT

What are your ideas for shaping AI to serve the public good?

Global report of the consultations among citizens and civil society actors ahead of the AI Action Summit

December, 2024

SciencesPo
Paris School of International Affairs

AI & Society Institute

Fondation Abeona
Dauphine | PSL
ENS | PSL

THE FUTURE SOCIETY

conseil national du numérique

MAKE.ORG

Summary

INTRODUCTION

Context: Summary and Key Findings

SciencesPo
PARIS SCHOOL OF INTERNATIONAL AFFAIRS

AI ACTION SUMMIT

Towards an inclusive AI governance process

The 2025 AI Action Summit has the potential to be a stepping stone for AI governance, akin to the 1992 Rio Conference for environmental policy:

- A historic high-level meeting to address civilizational challenges with bold commitments;
- A forum for civil society, industry leaders, researchers, and policymakers to design, discuss, and adopt relevant and applicable solutions;
- The foundation for a recurring engagement process for global, multistakeholder AI governance, similar to the COPs after Rio.

To amplify the Summit's inclusivity, legitimacy, and impact, a coalition including the AI and Society Institute (ENS-PSL), CNNum, The Future Society, Make.org, and Sciences Po's Tech & Global Affairs Innovation Hub, with the support of AXA, Capgemini Invent and Fathom, coordinated two open consultations between Sept. and Dec. 2024.

Thousands of contributions were received from experts, civil society organizations, and citizens from around the globe, thus enriching the Summit's upcoming agenda, discussions, and outcomes.

While our increasingly fragmented and polarized digital ecosystems are reminders of the shortcomings of the digital tech policy over the past decades, the voices of 10 000+ citizens and 200+ experts clearly express a strong demand for robust governance of emerging AI technologies, at the national and international level.

This report outlines priorities and actionable recommendations to protect citizens' rights, ensure fairness, and maximize the positive impact of AI systems, whilst safeguarding against their risks. Participants call for leveraging AI to address current societal issues such as medical research, climate change, disinformation… and push for inclusive, bottom-up governance where they can remain engaged, from policy design to ongoing evaluation.

This consultation marks only the beginning. It demonstrates that an inclusive and decentralized approach can transcend the diverging private and national interests, paving the way for inclusive multistakeholder AI governance, at the service of the greater good.

A broad basis to map the public opinion, increase inclusivity and build legitimacy for the Summit.

Citizen consultation

from 18 Sept. to 4 Nov. 2024

11,661 participants	**575** proposals	**121,325** votes

Expert & CSOs consultation

from 18 Sept. to 15 Nov. 2024

200 + experts & civil society orgs	**5** continents represented	**15** key deliverables

Key takeaways from the open consultation

1. **The public opinion demonstrates a sophisticated understanding of AI**

   Participants are numerous and demonstrate nuanced and diverse opinions of AI's potential and risks. Despite the technical nature of the matter, the level of awareness validates the importance of involving the public and civil society in the governance of AI.

2. **Participants reject any form of AI solutionism and uncontrolled deployments**

   Participants call for robust governance frameworks, both at local and international levels, to safeguard their rights and protect human agency. They are divided about unchecked deployments of AI systems and reject the idea of leaving key decisions to private companies.

3. **They want resources to learn, experiment, and progressively get to grips with the technology**

   The consultation shows a widespread demand for resources to help people better understand and engage with AI. This includes generalizing AI training across all levels of education and offering widely accessible professional development opportunities for people to harness the potential of AI but also to better comprehend -and control- their uses.

4. **And collectively use AI to answer concrete problems where it has added-value**

   The consultation underlines how participants prioritize safe deployments of accountable systems, addressing real-world problems. Sectors with particularly high support mostly involve specialized tools that work to enhance human expertise and institutional trust, such as health, public innovation, climate action, and the fight against disinformation.

Complementarity between citizens, experts and civil society organizations (CSOs)

1. **Experts and CSOs agree on the need for stronger multistakeholder AI governance**
   The contributions unanimously emphasize the urgent need for robust AI governance frameworks to regulate the activities of profit-driven companies. CSOs press for their involvement in the elaboration of governance mechanisms, their implementation, and the ongoing evaluation of regulation.

2. **Governance must set standards for auditable, fair, environmental-friendly, and privacy-respecting AI models**
   Participants advocate for the establishment of global frameworks to evaluate and compare AI tools on fairness, explainability, environmental footprints, transparency, and privacy. Standards must safeguard existing rights, promote accountability, and guide public and private procurement and investment.

3. **Respondents widen the scope to include safety issues, systemic risks and long-term threats**
   While the public opinion is more sensitive to immediate risks and harms, CSOs and experts remain vigilant towards AI safety and existential risks. They call for better-funded and coordinated AI safety research, as well as the enforcement of thresholds and other controls on large-scale models.

4. **And back proposals promoting equitable AI and limiting systemic inequalities amplification**
   The consultation highlight a common concern around the fair distribution of the benefits of AI. CSOs and experts jointly call for the adoption of mechanisms that reduce biases, enhance representation, and ensure AI's potential can be harnessed effectively and equitably by all, regardless of their origin, location, gender, or income.

Top priorities for the AI Action Summit

Lay the foundation for strong, global, multistakeholder AI governance

Call on global leaders to regulate development and deployment of AI tools and put a check on private interests. Design internationally coordinated laws, regulations, standards, and other mechanisms that are elaborated, implemented, and evaluated in collaboration with civil society organizations.

Target stream: #Global AI Governance

Suggested Deliverables

Launch the International Scientific Panel on AI Risks

Establish Mechanisms for Civil Society Organizations (CSOs) and Global South Inclusion

Top priorities for the AI Action Summit

Ensure access to qualitative AI education and training for everyone

Empower all individuals–regardless of origin, gender, age, or income–to build AI literacy, develop their skills, to safely and fairly reap the benefits of AI. Accessible education is essential for citizens to evaluate tools and use cases, assert their rights and actively participate in governance.

Target stream: #Future of Work

Suggested Deliverables

Launch a Global AI Education and Critical Thinking Initiative

Start an International Task Force on AI and Labor Market Disruption

Top priorities for the AI Action Summit

Establish shared standards for safe, responsible AI

Expand existing safety evaluation mechanisms to monitor systemic risks and adopt common benchmarks for transparency, fairness, explainability, and accountability. These standards should define baselines, promote good practices, equip regulators, and inform investments and procurement in the public and private sectors.

Target stream: #Trust in AI

Suggested Deliverables:

Establish and Publish Thresholds for AI Oversight

Launch the AI Corporation Commitments Report Card

Top priorities for the AI Action Summit

Ensure AI's Environmental Benefits Outweigh Its Costs

Establish standardized, verifiable benchmarks to measure the environmental footprints for AI models (including greenhouse gas emissions, water usage, minerals consumption), across the entire AI value chain. Promote AI use cases that drive sustainability, reduce ecological harm, and mitigate climate change effects.

Target stream: #Innovation & Culture

Suggested Deliverables:

Create mandatory, auditable model efficiency standards

Set up a GreenAI Leaderboard

Top priorities for the AI Action Summit

Prioritize AI solutions addressing existing issues, protecting our rights

Acknowledging the risks of uncontrolled AI development and indiscriminate deployment, the Summit must accelerate the identification of real-world problems where the added value of AI is clear and equitably shared. AI solutions must safeguard existing rights and enhance public trust to ensure responsible and meaningful innovation.

Target stream #Public Interest AI

Suggested Deliverables:

Launch a Global Charter for Public Interest AI

Launch the AI Commons Initiative to Empower Citizens in AI Design

PART 1

Findings from the citizen consultation

MAKE. ORG

AI ACTION SUMMIT

AI ACTION SUMMIT

What are your ideas for shaping AI to serve the public good?

Report of the citizen consultation
December, 2024

SciencesPo
Paris School of International Affairs

AI & Society Institute

Fondation Abeona
Dauphine | PSL
ENS | PSL

THE FUTURE SOCIETY

conseil national du numérique

MAKE.ORG

with the support of

AXA

Capgemini invent

Fathom

1st part

Consultation overview

MAKE.ORG

AI ACTION SUMMIT

Why this consultation?

At the initiative of the President of the Republic, France will host the International Summit for AI Action in February 2025. Artificial Intelligence is transforming jobs, health, culture, economies, and numerous other sectors worldwide. This Summit aims to address this critical topic with experts and stakeholders.

The online consultation sought to broadly engage citizens and civil society, gathering ideas on how to make artificial intelligence an opportunity for all while collectively preventing inappropriate or abusive uses of this technology. Participants were invited to answer the question:

"What are your ideas for shaping AI to serve the public good?"

The results of this consultation will be analysed and submitted to the AI Summit's working groups.

from 16.09.2024 to 04.11.2024

Key figures from the consultation

What are your ideas for shaping AI to serve the public good?

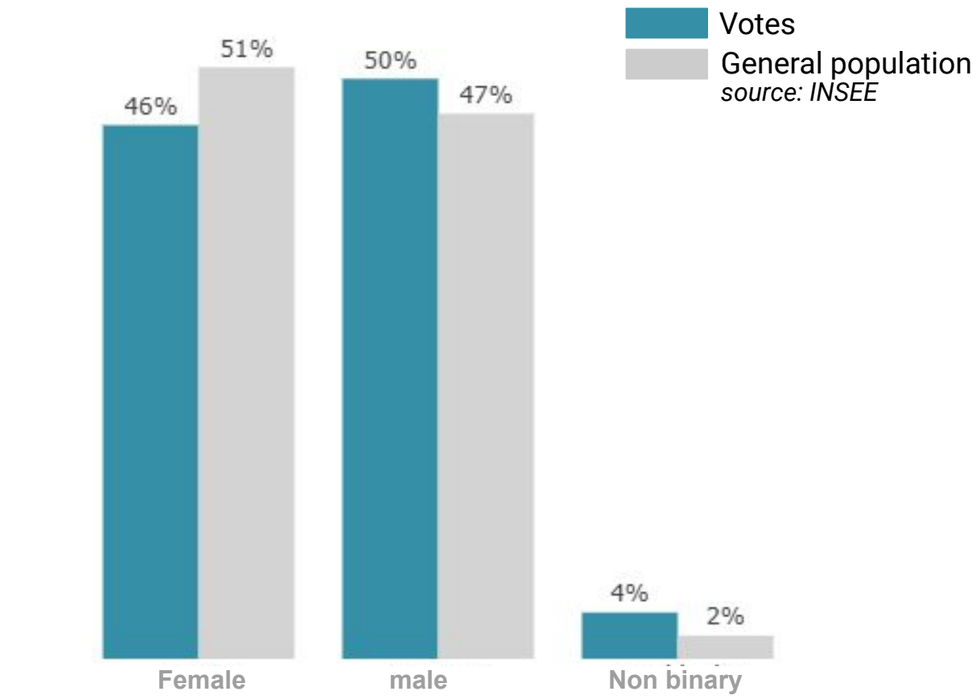
11,661
participants

575
proposals submitted

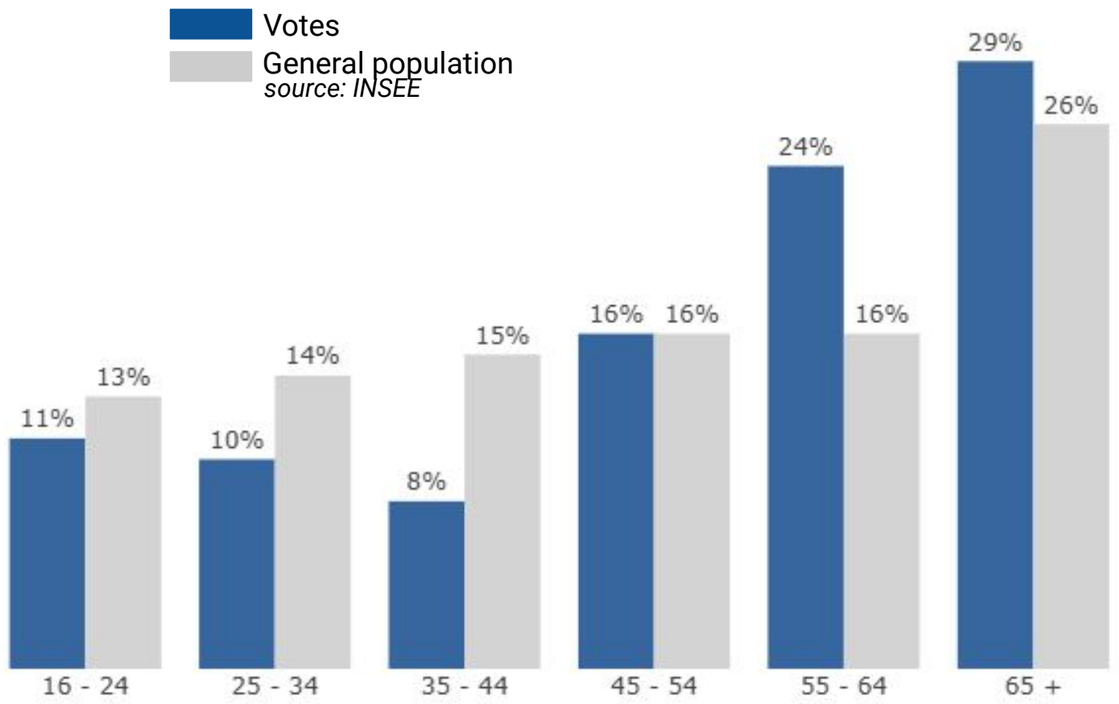
121,325
votes

Breakdown of Participation in the Consultation by Gender, Age, and Status

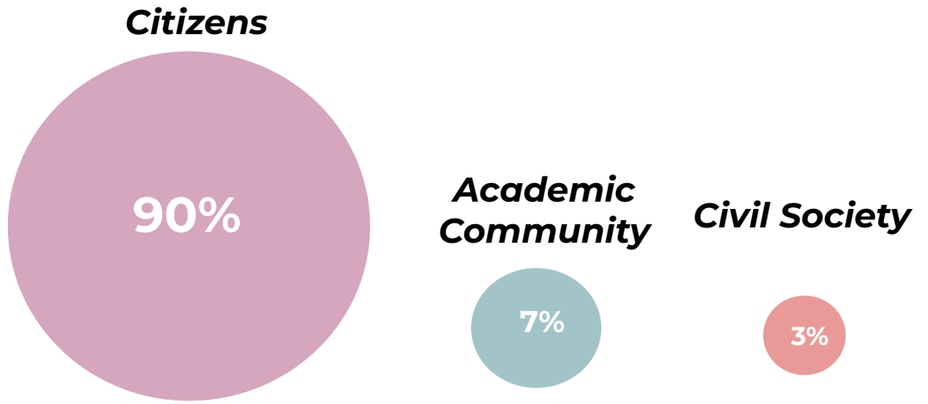
Participation by Gender



Participation by Age



Participation by Status

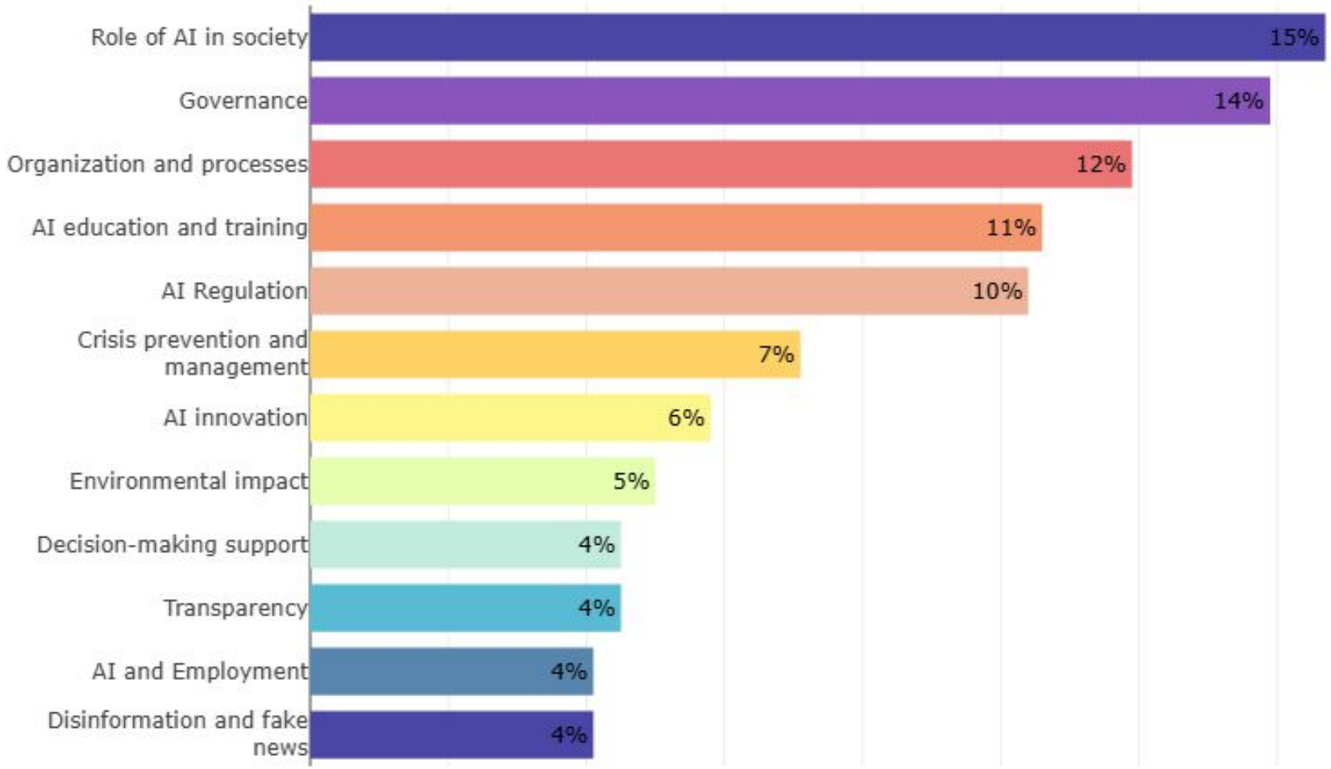


Main themes of the consultation

What citizens are talking about

% of 502 validated proposals*

**The sum of the percentages is greater than 100% because some proposals fall under more than one theme.*



This graph does not take citizens' votes into account, only the number of proposals.

2nd part

Popular ideas and controversial ideas

MAKE.
ORG

AI ACTION
SUMMIT

Methodology

575 PROPOSALS SUBMITTED TO THE CONSULTATION
502 VALIDATED PROPOSALS

Consensus zone

227 proposals

More than 60% of votes in favour

Controversy zone

253 proposals

> 60% of votes against
Rejected

> 40% neutral votes
Low-frequency

17

6

Fewer than 60% of votes in favour
More than 15% of votes against

To develop the ideas, proposals on the boundaries of each zone were filtered (over 63% “in favour" votes in the consensus zone; fewer than 55% “in favour" votes and more than 20% "against" votes in the controversy zone) to isolate the most significantly supported or controversial ones.

Qualitative analysis by grouping together proposals that convey similar ideas

15 Popular ideas

9 Controversial ideas

15 popular ideas, 9 controversial ideas

The most popular ideas (> 5 popular proposals)

The most controversial ideas (> 2 controversial proposals)

AI Education

- Train the population in the ethical use of AI and raise awareness of its biases.
- Integrate AI into educational curricula, both as a subject and as a tool.
- Improve accessibility to AI for the widest possible audience.

Governance and Democracy

- Monitor the expansion of AI and better define its role in society.
- Harmonise ethics and global governance of AI.
- Leverage AI to safeguard democracies.
- Strengthen legal frameworks to better regulate AI usage.
- Utilise AI to guide public policies and democratic processes.

Transparency and Trust

- Ensure easy identification of AI-generated content.
- Guarantee the security of personal data.
- Expand funding for research on AI biases.

Public Interest

- Develop AI for health diagnostics.
- Gradually use AI to optimise public services.
- Address public interest issues through AI.
- Streamline public administration management with AI.
- Simplify access to justice and its functioning.
- Facilitate administrative procedures with the help of AI.

Managing the Impact of AI on Society

- Limit AI to specific uses and sectors.
- Prevent and manage the impact of AI on work and employment.
- Ensure the protection of intellectual property, cultural and artistic production.
- Stop the use of AI.
- Generalise professional transition programmes to address the changes brought by AI.

AI and the Environment

- Limit the environmental impact of AI.
- Leverage AI to anticipate natural crises.

01.

AI Education

MAKE.
ORG

AI ACTION
SUMMIT

Popular idea

Train the population in the ethical use of AI and raise awareness of its biases

17 proposals

What the majority of citizens agree on

Training a large number of citizens, especially the younger generation, in AI with the aim to foster responsible use: cultivating a critical mindset towards this technology, teaching its associated risks and biases.

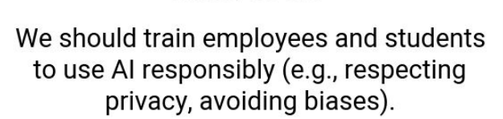
Understanding how AI works in order to better guard against these biases and develop an ethical approach to the tool.

Prioritising training for students (for example, with mandatory AI introduction courses, demonstrations, and workshops in schools or universities), workers, and also continuous training for data scientists and even decision-makers.

❤️ Popular proposal examples:

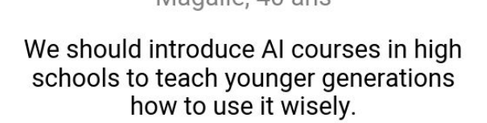
Anne, 45 ans

We should train employees and students to use AI responsibly (e.g., respecting privacy, avoiding biases).



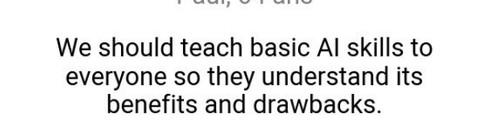
Magalie, 40 ans

We should introduce AI courses in high schools to teach younger generations how to use it wisely.



Paul, 64 ans

We should teach basic AI skills to everyone so they understand its benefits and drawbacks.

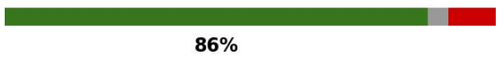


Benjamin, 39 ans

We should organise community hubs, like libraries, to discuss AI-related questions and concerns.

Nicole, 65 ans

We should prioritise informing and educating people to maintain critical thinking about this new technology.



Jules, 20 ans

We should include AI in school curricula to teach its workings and discern when it is truly useful.

% votes "in favour" % “neutral” votes % votes "against"

Controversial idea

Integrate AI into educational curricula, both as a subject and as a tool

21 proposals

What citizens are divided on

Creating AI platforms for students to help them revise, do exercises, or learn.

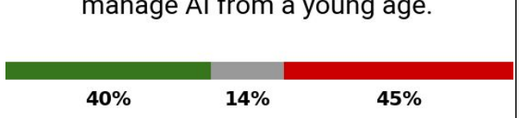
Relying on AI to train teachers, assist in preparing lessons, and assess students: essentially, any intervention that would "replace" teachers.

Directly integrating AI courses into school curricula, or even offering an AI certification within the programme.

⚡ Controversial proposal examples:

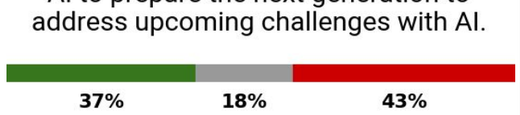
The

We should revise educational programmes so children learn how to manage AI from a young age.



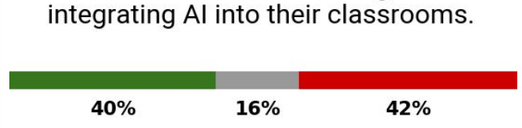
Bhupesh

We should establish a professional training and certification school for AI to prepare the next generation to address upcoming challenges with AI.



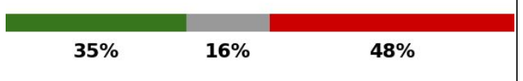
Laurent, 33 ans

We should ensure secondary school teachers receive training on integrating AI into their classrooms.



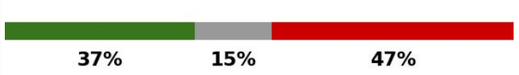
Jean, 40 ans

We should provide "extra teacher" AI tools to explain course concepts differently or generate exercises for students.



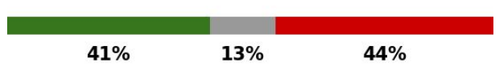
Camille, 36 ans

We should use AI to help teachers analyse student progress in real time and adjust their teaching strategies.



Jean, 40 ans

We should allow the creation of AI tools to help students study by generating exercises and explaining course concepts.



% votes "in favour" % “neutral” votes % votes "against"

Controversial idea

Improve accessibility to AI for the widest possible audience

10 proposals

What citizens are divided on

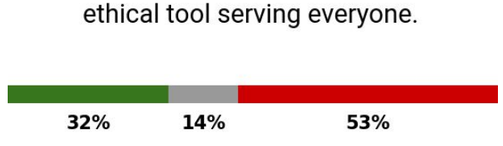
Making AI free to ensure the widest possible access for the entire population, especially for the most disadvantaged groups to benefit from it.

Citizens are divided on whether to invest resources for this purpose.

⚡ Controversial proposal examples:

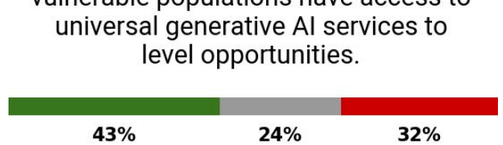
Claire, 75 ans

We should build trust in AI and make it an ethical tool serving everyone.



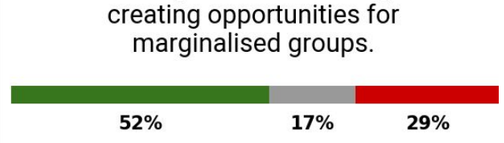
Jean, 41 ans

We should ensure that the most vulnerable populations have access to universal generative AI services to level opportunities.



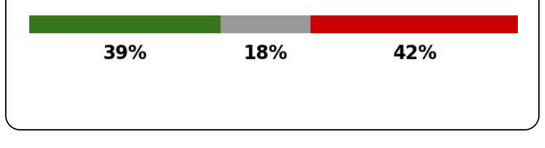
Didier, 54 ans

We should use AI to serve communities by facilitating distance learning and creating opportunities for marginalised groups.



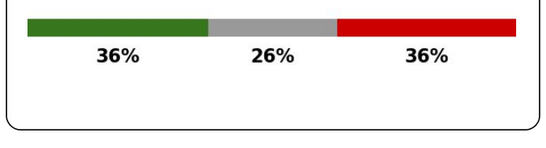
The

We should make AI tools free or as accessible as possible to the general public.



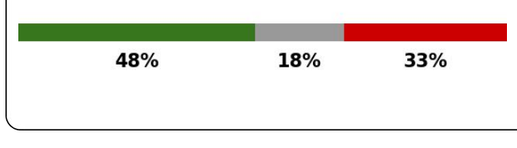
Jean, 41 ans

We should guarantee that everyone worldwide has access to developed AI services.



David, 46 ans

We should ensure everyone learns to use AI, providing both time for training and resources for accessibility.

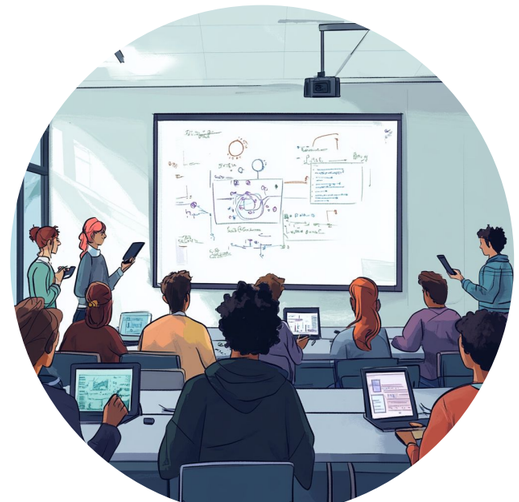


% votes "in favour" % “neutral” votes % votes "against"

Educating about AI technologies

"The question is no longer whether to make room for AI in education – it has already made its way in – but how to support ~~the~~ ongoing developments and address the challenges of educating by and with AI."
Senate report on AI and education, 2024

Thanks to a text-based interface accessible to every citizen, generative artificial intelligence (GenAI) has become the fastest and most widely adopted technology in modern history. It took ChatGPT only 5 days after its release to reach one million users! However, this easy and widespread use by the general public does not eliminate the challenges related to training, awareness, and education posed by this new technology. Today, between 80 and 90% of young people aged 13 to 21 regularly use generative AI, mainly for school tasks such as reviewing courses, summarising, translating, and creating texts. The reasons for this enthusiasm include the AI's infinite patience, its benevolence, and its ability to meet students at their level without judgement. Therefore, the question of teaching about AI and how to use it is becoming increasingly important, both for our education systems and for our societies in general, in order to ensure its reliable and ethical use and to promote equal access for all.



However, France is lagging behind in AI education, especially in terms of teacher training: only 8% of higher education instructors regularly use AI, while 65% do not use it at all. France is not alone in this regard: globally, only seven countries have AI training programmes or frameworks for teachers: China, Spain, Finland, Georgia, Qatar, Thailand, and Turkey. This disparity highlights the urgent need for teacher training and the integration of AI into educational practices, so that educators can prepare future citizens to use this technology while addressing its risks and benefits. The costs associated with such efforts in terms of time (training professionals) and resources (developing tools) are considerable.

Generative AI could represent a major technological breakthrough in the way we teach, practice, learn, evaluate work, and organise school life. For students, AI facilitates research, supports writing tasks, enables the creation of exercises tailored to individual needs, and encourages mentorship. For teachers, AI can be used to create specialised content adapted to students' needs, prepare engaging versions of their lessons (such as video materials, podcasts, etc.), and automate administrative tasks (e.g., assistance with filling out forms). Finally, for educational institutions, AI can help streamline mundane tasks, optimise timetable management, and equip schools to combat plagiarism, among other applications.

However, the rapid development of artificial intelligence, particularly generative AI, and its integration into educational curricula have sparked numerous debates within the educational community. The integration of AI in the classroom and its framework of use are significant points of discussion, and scientific evidence regarding the pedagogical benefits of AI is being closely examined, with concerns about the erosion of fundamental skills such as reading, writing, critical thinking, and self-assessment. Similarly, there are concerns about the transformation of the teaching profession, particularly the risk of "dispossession." Data security and the sovereignty of the French education system are also major issues, especially when AI models used are not of French origin.

Therefore, an ethical and trustworthy framework for AI in education must be developed to guard against biases (for instance, if student evaluations are conducted by AI), to determine the appropriate age for introducing AI in schools, and to establish the forms in which this new competency should be taught in order to genuinely benefit both teachers and students.

02.

Governance and Democracy

MAKE. ORG

AI ACTION SUMMIT

Popular idea

Monitor the expansion of AI and better define its role in society

34 proposals

What the majority of citizens agree on

Keeping AI under human control to prevent abuses and potential loss of control, while preserving the human aspect of social and learning interactions and maintaining their spontaneity.

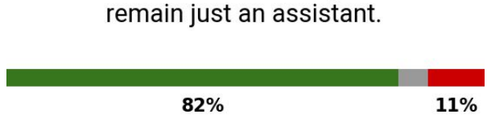
Ensuring that AI development is monitored and that it remains a tool for automation, without interfering in human decision-making, and tempering excessive belief in AI's virtues.

The fear of AI replacing humans is emerging, highlighting the need to ensure this technology serves humans, not the other way around: "control AI before it controls us."

❤️ Popular proposal examples:

Christophe, 58 ans

We should ensure governments stop this race toward consumerism; AI should remain just an assistant.

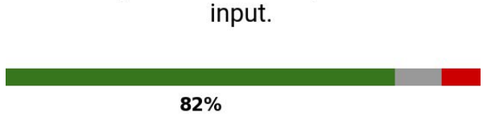


Simon, 59 ans

We should not overestimate AI's promises to solve complex problems when deciding which actions to take.

Emmanuelle

We should ensure AI remains a servant to humanity and does not replace human input.

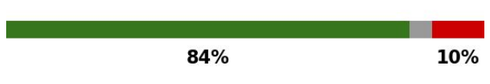


Claire, 75 ans

We should use AI as a decision-making aid but never blindly follow its recommendations.

Vr

We should subject all AI developments to human oversight to prevent its autonomy and preserve our own.



Fabienne

We should reorient priorities to place humans at the centre of social interactions.

% votes "in favour" % "neutral" votes % votes "against"

Popular idea

Harmonise ethics and global governance of AI

20 proposals

What the majority of citizens agree on

Creating an ethical committee at a global or European level, capable of establishing a solid and uniform ethical framework (charters, AI usage oath, individual responsibilities) to prevent AI abuses, as well as technical standards or even regulations, similar to those for human rights or genetics.

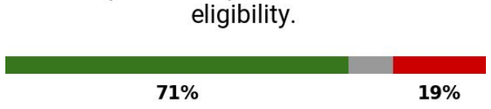
Producing reports on AI, similar to the IPCC's reports on climate change.

A commitment involving governments, civil society, academia, and also the AI giants is also proposed.

❤️ Popular proposal examples:

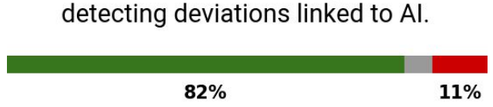
Sine, 58 ans

We should establish an international code of ethics for AI use, with compliance required for contract eligibility.



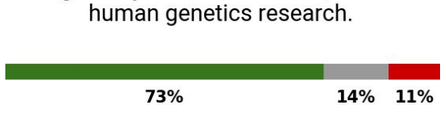
Patricia

We should establish ethical AI committees in every organisation, staffed with humans capable of detecting deviations linked to AI.



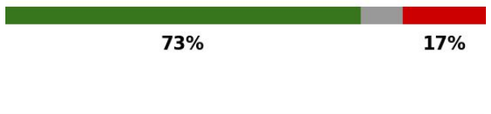
Bibi, 50 ans

We should pursue an international regulatory framework for AI, akin to human genetics research.



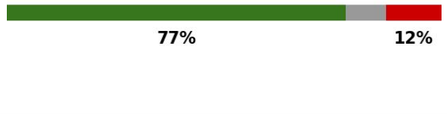
Axa

We should establish ethical governance and regulations for AI development and use in line with societal values and human rights.



Lucie, 39 ans

We should ensure AI giants honour their ethical commitments made at global summits.



Marie, 24 ans

We should publish annual reports, akin to those from the IPCC, on AI risks and opportunities to guide global decisions.

% votes "in favour" % "neutral" votes % votes "against"

Popular idea

Leverage AI to safeguard democracies

13 proposals

What the majority of citizens agree on

Strongly combating the proliferation of AI-generated fake news that destabilises democratic life and elections.

Limiting foreign political intrusions and interference in debates. AI could, in this regard, be used to detect such manipulations.

Remaining vigilant regarding the use of AI in managing individual freedoms (e.g., in China or facial recognition), particularly by strengthening citizen oversight.

Establishing European AI standards to ensure the protection of human rights.

❤️ Popular proposal examples:

Lena, 60 ans

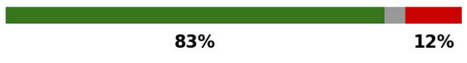
We should demand ethical AI and guarantee its democratic oversight through regular consultations.

Nicolas

We should use AI to fight the information war by detecting fake news and manipulated media.

Anne, 59 ans

We should avoid using AI in ways that harm individual freedoms, as seen in China; security does not justify every use.

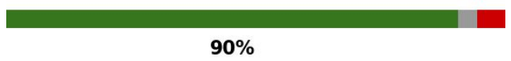


S. Thierry, 60 ans

We should ensure AI is not used for political purposes.

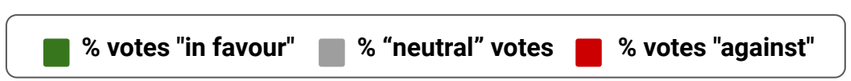
Axa

We should protect our democracies by ensuring that AI is not used to fuel disinformation and misinformation.



Marthe, 26 ans

We should establish international AI standards, inspired by European law, to protect global citizens' rights.



Popular idea

Strengthen legal frameworks to better regulate AI usage

12 proposals

What the majority of citizens agree on

Creating new, specific, and harmonised regulations at a European level, such as a dedicated government agency, to prevent malicious uses and protect vulnerable individuals. Citizen oversight is also considered.

Implementing AI moderation for platforms, similar to the CE marking system, coupled with a reporting system.

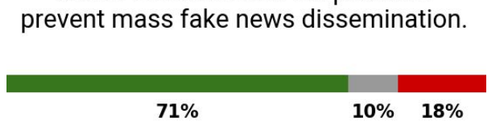
❤️ Popular proposal examples:

Robert, 70 ans

We should implement strict legislation to penalise malicious uses of AI.

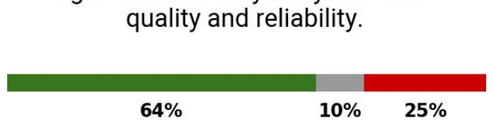
Robinson, 38 ans

We should develop moderation tools for social networks and the public to prevent mass fake news dissemination.



Marjorie

We should establish public regulatory agencies to certify AI systems for quality and reliability.



Bernard, 70 ans

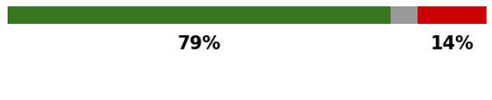
We should develop AI moderators to monitor its applications.

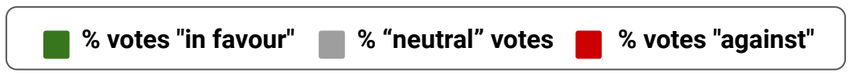
Solenne, 31 ans

We should require companies to regularly evaluate AI systems and mitigate detected risks.

Morgane, 27 ans

We should strengthen legal frameworks around AI to prevent abuses (e.g., deepfakes, copyright infringement) at a European level.





Controversial idea

Utilise AI to guide public policies and democratic processes

19 proposals

What citizens are divided on

Relying on data to identify specific needs in local policies and, more broadly, in the management of certain public policies: health, public transport, employment, etc.

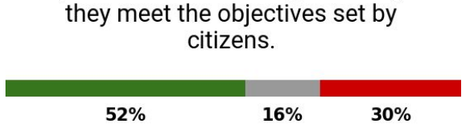
Improving the functioning of representative institutions (e.g., ministries, European institutions) and their elected officials through AI.

Using AI in citizen representation or voting, or to stimulate participatory democracy.

⚡ Controversial proposal examples:

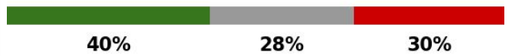
Dominique, 63 ans

We should develop AI to monitor political accountability, ensuring they meet the objectives set by citizens.



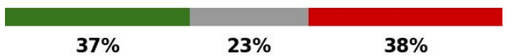
Marjorie

We should establish citizen AI assemblies in major metropolitan areas.



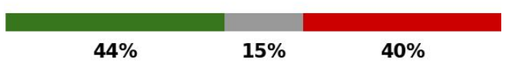
Paul, 64 ans

We should create a European AI tool enabling citizens to understand, communicate, and act at an EU level.



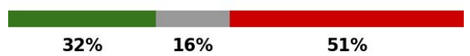
Marthe, 32 ans

We should use AI to map social needs and adjust local policies accordingly.



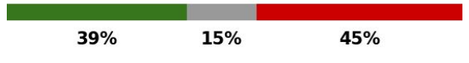
Jessica, 37 ans

We should create an AI-augmented democracy platform to engage citizens in voting, contributing, and understanding public issues.



Gwenaelle, 43 ans

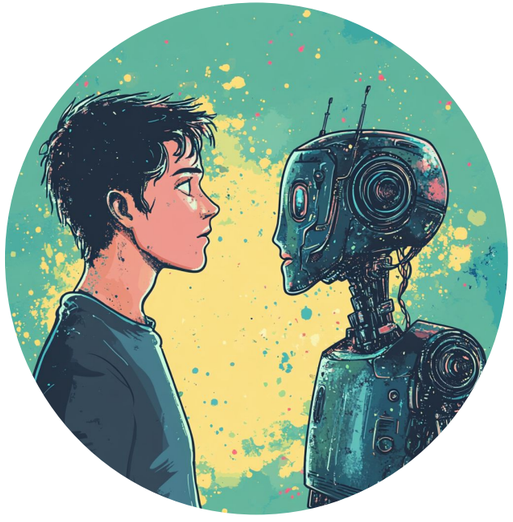
We should use AI to predict epidemics so health ministers can act promptly, as with COVID-19.



% votes "in favour" % “neutral” votes % votes "against"

Regulate to unlock the full potential of AI

"We must regulate AI to ensure it is used in ways that benefit society. Without proper governance, the risks to democracy and human rights are too great."
Stuart Russell, co-author of *Artificial Intelligence: A Modern Approach*



The adoption of artificial intelligence is expanding rapidly, particularly with the rise of generative AI tools that allow anyone to create text, images, videos, sounds, and more. This technology offers boundless opportunities to modernise institutions and strengthen connections between citizens, their representatives, and public services. For instance, AI tools can present public debates, new laws, or the conditions for obtaining aid or subsidies in an accessible and educational way. The creation of collaborative platforms – run and enhanced using generative AI – can also foster civic engagement and participation in decision-making by analysing public opinions and helping to design inclusive policies.

However, the absence of regulation in AI deployment poses significant risks to democracy. Numerous scandals, such as the "Cambridge Analytica" controversy, revealed the dangers of mass manipulation, particularly through the spread of disinformation (fake news, deep fakes) and mass surveillance technologies. Furthermore, AI can amplify existing biases and discrimination when not developed within a rigorous framework. These concerns underscore the urgent need for robust technical and ethical oversight of AI use and the establishment of broad, inclusive governance, given that AI is a global technology. Efforts towards regulation are beginning to emerge, such as the European Union's AI Act, alongside the development of ethical charters and frameworks aimed at maximising AI's benefits while mitigating its risks.

AI technologies themselves can play a role in regulating their use. For instance, the concept of "human in the loop" integrates regular human oversight into AI systems, creating hybrid models that maintain human supervision. Similarly, additional layers of AI can act as regulatory mechanisms – essentially AI controlling AI (e.g., large language models as judges). It is also possible to constrain AI models by defining their scope of validity and the datasets on which they were trained, ensuring prediction compliance. Automated auditing models are another promising AI oversight approach.

The European Union's AI Act, designed to regulate the use of artificial intelligence, introduces a classification of AI systems based on their potential risk, ranging from unacceptable to minimal. High-risk systems will be required to meet strict standards for transparency, safety, and human oversight. The aim is to foster ethical and reliable AI while safeguarding citizens' fundamental rights. This regulation is a trailblazer and could serve as a global benchmark for AI governance. The AI Commission has also proposed global mechanisms, including the creation of a World AI Organisation to evaluate and oversee AI systems, an International AI Fund to promote initiatives serving public interest, and a "1% AI Solidarity Mechanism" to support developing countries. However, achieving such governance poses significant challenges due to the rapid pace of technological development and AI's global reach. Regulations often struggle to keep up with innovations, leaving policymakers perpetually behind. Additionally, the costs associated with monitoring and enforcing AI regulations are substantial, presenting a considerable hurdle for many nations.

03.

Managing the Impact of AI on Society

MAKE.
ORG

AI ACTION
SUMMIT

Popular idea

Limit AI to specific uses and sectors

27 proposals

What the majority of citizens agree on

Collectively determining – particularly through debates – where the development of AI should be accelerated or not. At a minimum, carefully assessing its utility and impacts before applying it indiscriminately in these areas.

Restricting AI use in specific fields: in education, recruitment processes, the arms sector, etc. Its use, however, is more widely accepted in sectors such as medical research.

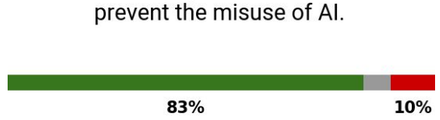
❤️ Popular proposal examples:

Jean, 40 ans

We should set risk thresholds, beyond which AI development or deployment must be paused.

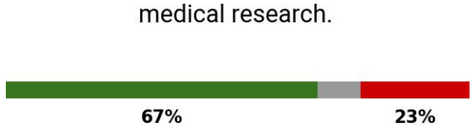
Axa

We should take proactive measures to prevent the misuse of AI.



Marie-Louise

We should limit AI use to scientific and medical research.



Natacha

We should not allow AI in recruitment processes.

Alice, 18 ans

We should see AI as a useful tool in many areas but never as the final result - only as an aid for professionals.

Simon

We should regulate AI use in schools.

% votes "in favour" % “neutral” votes % votes "against"

Popular idea

Prevent and manage the impact of AI on work and employment

8 proposals

What the majority of citizens agree on

Anticipating the evolution of AI and the risks of job displacement in certain sectors.

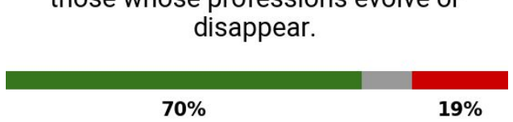
Preventing widespread job loss by implementing countermeasures and finding solutions for those who lose their positions.

Supporting transitions by redefining jobs to ensure real protection and avoid sudden replacement by AI.

❤️ Popular proposal examples:

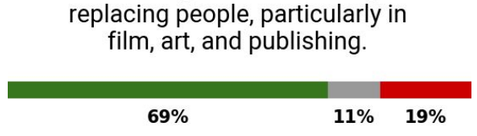
David, 46 ans

We should provide help with societal changes brought by AI, especially for those whose professions evolve or disappear.



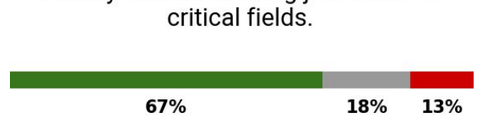
Lily, 19 ans

We should use AI to improve lives without diminishing human creativity or replacing people, particularly in film, art, and publishing.



Jessica, 37 ans

We should ensure automation benefits society without causing job losses in critical fields.



John

We should anticipate AI's impact on professions and develop behavioural skills to ensure human well-being.

Mathieu, 27 ans

We should implement safeguards to ensure AI cannot replace us in critical technologies.

Ele, 17 ans

We should prevent AI from replacing humans and their jobs. Machines don't need wages to live - people do.

% votes "in favour" % “neutral” votes % votes "against"

Popular idea

Ensure the protection of intellectual property, cultural and artistic production

6 proposals

What the majority of citizens agree on

Granting AI a role in art requires an international regulatory framework to protect artists from the exploitation of their work and ensure fair compensation.

AI creations should belong to the public domain, and their use should be taxed to prevent unfair competition. There is also a need to adapt copyright laws to this new reality, ensuring fair compensation for artists whose data may have been exploited without consent.

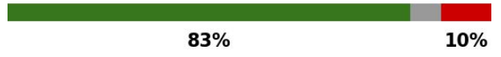
❤️ Popular proposal examples:

Ioan, 27 ans

We should create new copyright systems to fairly compensate artists whose work trains AI models.

Dominick, 56 ans

We should create an international regulatory framework to allow artists to protect their voices and images from AI-driven data theft.



Jessica, 37 ans

We should compensate artists whose works have been exploited by AI without consent.

Une Artiste Comme Les Autres, 19 ans

We should ban AI in the arts when it involves stealing works to create others without meaningful contribution.

Pierre-Emmanuel, 41 ans

We should legislate so AI creations fall into the public domain, avoiding private ownership of their output.

Thibaut, 34 ans

We should tax AI-created artistic content to counter unfair competition with human creators.

% votes "in favour" % “neutral” votes % votes "against"

Controversial idea

Generalise professional transition programmes to address the changes brought by AI

14 proposals

What citizens are divided on

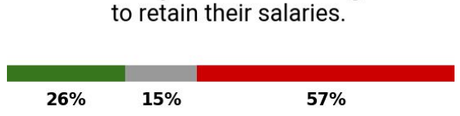
Forcing transformation and job transitions in companies because of AI, requiring companies to train their staff, or generalising training and job replacements.

The general idea of accepting that certain jobs will disappear and that AI is a positive solution, creating new roles, with inevitable career transitions.

⚡ Controversial proposal examples:

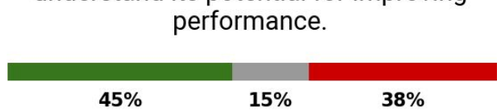
Antoine, 50 ans

We should use AI to replace those who dislike their jobs while allowing them to retain their salaries.



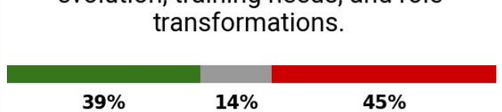
Axa

We should prepare people for present and future jobs by teaching them to use AI and understand its potential for improving performance.



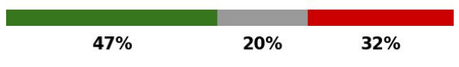
M

We should require companies to plan for AI integration, detailing job evolution, training needs, and role transformations.



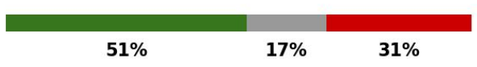
Olivier, 28 ans

We should generalize AI "side school" training to help everyone anticipate technological shifts in their professions.



Jean, 40 ans

We should raise awareness about how some jobs will disappear with AI, but others will emerge.



Alexis, 54 ans

We should support career transitions to prepare workers for the AI era.

% votes "in favour" % “neutral” votes % votes "against"

Controversial idea

Stop the use of AI.

14 proposals

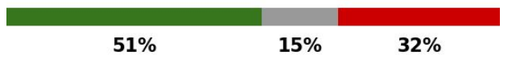
What citizens are divided on

A complete ban on the use of AI, which is viewed as a danger or nuisance, potentially leading to both social and human regression, as well as environmental harm.

⚡ Controversial proposal examples:

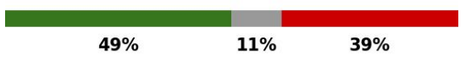
Jean-Jacques, 53 ans

We should avoid using AI due to its environmental impact, such as the energy consumption of data servers.



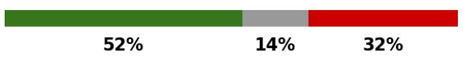
Maryline, 67 ans

We should stop using AI entirely, given the harmful and perverse applications we already see.



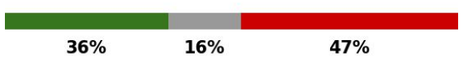
Sylvaine, 58 ans

We should ban AI entirely; it only enriches the wealthy and is a regression rather than progress.



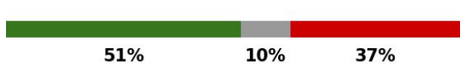
Jean Claude, 82 ans

We should simply eliminate AI altogether.



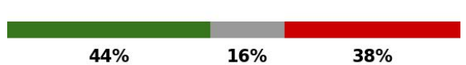
André, 75 ans

We should ban AI in the public sector and restrict its use to professional fields where it is genuinely needed.



Gégé

We should completely ban AI, as it is a destructive technology that eliminates jobs and creates misery.



% votes "in favour" % “neutral” votes % votes "against"

Driving change at a societal level.

"Artificial intelligence is the new electricity. It will transform every sector of the economy."
Andrew Ng, professor at Stanford University and co-founder of Coursera.

Researchers (Stanford, MIT) estimate that artificial intelligence could lead to productivity gains of 14% to 35%, depending on the nature of the activity, and Chris Pissarides, Nobel laureate in economics, believes that we could even easily transition to a four-day workweek. All sectors of our society are affected, and they will be increasingly impacted in the future, given the immense potential of AI. However, as noted by the French Government's AI Commission, "AI should neither provoke excessive pessimism nor excessive optimism: we anticipate neither mass unemployment nor automatic acceleration of growth. In the coming years, AI will not replace humans, nor will it be the solution to all the challenges of our time. We must neither overestimate its short-term impact nor underestimate its long-term effects." France and Europe, in particular, have all the assets needed to meet the challenges posed by AI in terms of societal change.

In particular, adapting professions is crucial and will help anticipate the evolution of the labour market and skills. Workers and entrepreneurs need to be prepared to evolve in an environment where AI plays an increasingly important role. This can be achieved through continuous training programmes and reskilling initiatives. It is also necessary to promote the development of transferable skills such as creativity, problem-solving, and project management, which are less likely to be automated. In France, the "Skills and Professions of the Future" (CMA) programme aims to adapt training to meet future needs. Similarly, adapting social protection systems should be considered to include safety nets for workers affected by automation. Conversely, it is also relevant to encourage the development of sectors where AI can create complementary jobs, such as AI system maintenance, data management, and cybersecurity.

Next, it is essential to invest in cutting-edge research and innovation to master the advancements driven by artificial intelligence and strengthen the competitiveness of our companies. The France 2030 plan aims to increase the capacity for AI training, with the emergence of 9 "AI clusters," universities, and prestigious schools that will become global hubs for training, research, and application. For its part, the AI Commission proposes the creation of a fund with €10 billion to finance the development of the AI ecosystem and the transformation of the French economic fabric, and, in the long term, to redirect French savings towards innovation. It is worth noting that strict regulation by governments, particularly concerning the protection of personal data, must be aligned with these goals to pursue innovation and competitiveness.

On the social front, some researchers highlight the risk of technological dependency, which could diminish individuals' critical thinking ability and lead to a loss of autonomy in decision-making. It is therefore essential to keep humans at the centre of our interactions to maintain human and social connections. Similarly, managing intellectual property and copyright is another important issue to address as a society, with the establishment of clear regulations to protect creations and innovations in a world where AI can autonomously generate content. Regular studies on the societal impact of AI, as well as platforms to monitor ethical guidelines within the framework of Trusted AI projects, could also be considered.

This adaptation is not without consequences, particularly financial ones, in a context of significant budgetary constraints. For example, the plan proposed by the AI Commission represents a public investment of €5 billion per year over five years.

04.

Transparency and Trust

MAKE.
ORG

AI ACTION
SUMMIT

Popular idea

Ensure easy identification of AI-generated content

11 proposals

What the majority of citizens agree on

Clearly identifying all multimedia content created by AI, through mentions and labels, similar to tagging systems; relying on reporting mechanisms.

Using AI to detect AI-generated content.

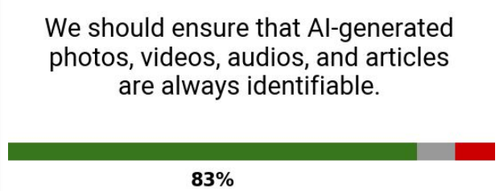
❤️ Popular proposal examples:

Adil, 45 ans

We should create ethical labels for AI solutions.

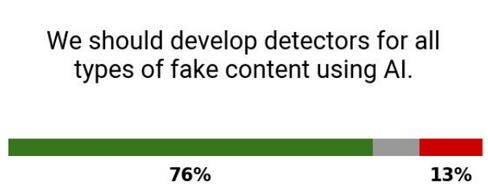
Patricia

We should ensure that AI-generated photos, videos, audios, and articles are always identifiable.



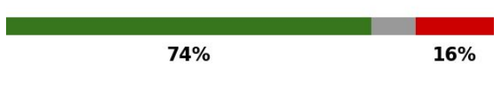
Evelyn, 33 ans

We should develop detectors for all types of fake content using AI.



Olivier

We should require clear labelling for all content partially or fully generated by AI, similar to food labelling.

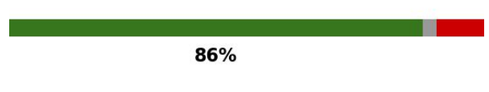


Lalie, 51 ans

We should notify citizens whenever an AI governs decisions, such as tax collection, benefit calculations, or phone calls.

Jean-Baptiste

We should ensure that all multimedia content created by AI is clearly identified as such, regardless of the platform.



% votes "in favour" % “neutral” votes % votes "against"

Popular idea

Guarantee the security of personal data

8 proposals

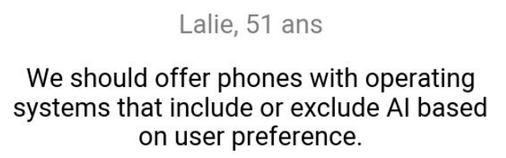
What the majority of citizens agree on

Monitoring and verifying the use of personal data by artificial intelligence and assessing privacy risks, such as those related to facial recognition. Each user should also be able to choose whether or not to activate AI-related features to preserve their autonomy.

❤️ Popular proposal examples:

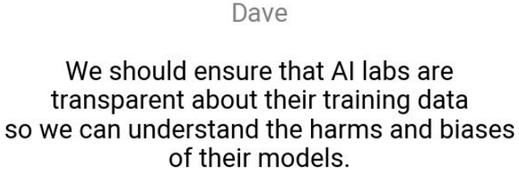
Lalie, 51 ans

We should offer phones with operating systems that include or exclude AI based on user preference.



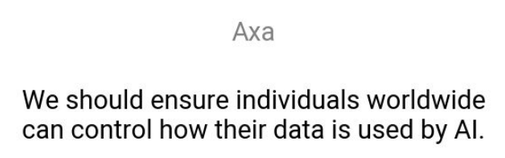
Dave

We should ensure that AI labs are transparent about their training data so we can understand the harms and biases of their models.



Axa

We should ensure individuals worldwide can control how their data is used by AI.



Susan, 70 ans

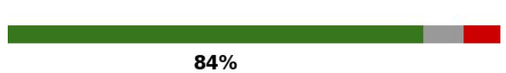
We should pay more attention to the misuse of data by major AI corporations and their "marketing" of nationalism in AI.

Philippe, 71 ans

We should require AI systems to assess their risks to privacy and human manipulation.

Marjorie

We should require conversational assistants to be transparent about the sources they use.



% votes "in favour" % “neutral” votes % votes "against"

Popular idea

Expand funding for research on AI biases

7 proposals

What the majority of citizens agree on

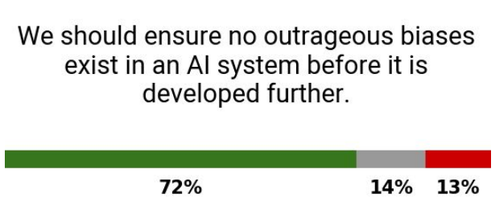
Funding research on AI for the public good and ensuring researchers have independent access to AI models to verify their quality and study biases.

Investing in the public regulation and transparent oversight of research, such as creating an international virtual lab to test AI security and better prevent its misuse.

❤️ Popular proposal examples:

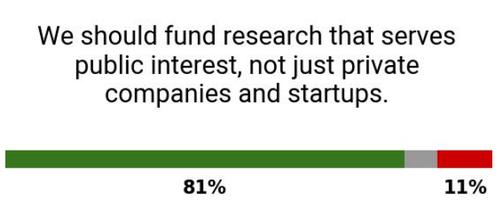
Jean, 41 ans

We should ensure no outrageous biases exist in an AI system before it is developed further.



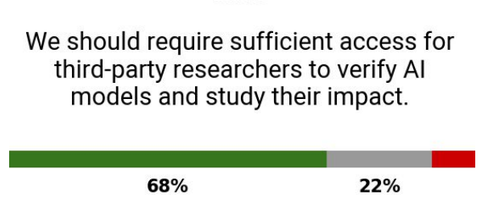
Marjorie

We should fund research that serves public interest, not just private companies and startups.



Dave

We should require sufficient access for third-party researchers to verify AI models and study their impact.



Dave

We should allow independent researchers access to verified models to conduct meaningful studies on potential harms.

Sandra, 39 ans

We should create an international virtual lab to measure and test AI safety.

Simon, 59 ans

We should increase public investment in independent AI research, oversight, and regulation for democratic control.

% votes "in favour" % “neutral” votes % votes "against"

Transparency is the first step towards accountability!

"The only way to ensure that AI is used responsibly is through transparency, accountability, and the involvement of a diverse group of people in its development."
Fei-Fei Li, researcher at Stanford University.

The rise of generative artificial intelligence, embodied by advanced technologies like GPT-5, is raising growing concerns about the potential flooding of the internet with AI-generated content. This context fuels a sense of opacity around AI solutions, with fears that they could perpetuate biases or discrimination. The transparency and auditability of AI systems thus becomes essential to ensure their operation is understandable and verifiable. Indeed, it is crucial to democratise these new technologies for the general public and provide transparency guarantees in order to foster trust and drive adoption, unlocking their full potential. Another key issue lies in information about the origin of these models and their ability to protect personal data, as it is also important to strengthen French and European technological sovereignty, balancing the influence of foreign actors (GAFAM, Russia, China).

To address these challenges, the new European regulation (AI Act) places particular emphasis on transparency and auditability. In terms of transparency, users must be informed when they interact with AI, and the functioning of the systems must be explainable. For auditability, detailed documentation and system traceability are required, enabling audits by competent authorities. Providers must also conduct compliance assessments before deploying high-risk systems. Human oversight is essential to allow intervention and contestation of AI decisions. Finally, mechanisms for monitoring and control by relevant authorities are foreseen to ensure the ongoing compliance of AI systems with the regulations.

In technological terms, explainable AI can be used to describe an AI model, its expected impact, and its potential biases. Interpretability models help characterise the outcomes of AI models in their decision-making processes, making them transparent. The development of ethical AI solutions - incorporating principles such as fairness, transparency, and accountability - and the integration of bias detection tools into AI models are also key to improving compliance and transparency. However, ethical debates, such as the question of responsibility in cases where significant decisions are made by AI, must be addressed collectively. Regarding data protection, federated learning aims to train AI models "remotely" while preserving the confidentiality of sensitive data. Finally, in the face of challenges like deep fakes, fake news, and identity theft, AI-based solutions are being developed to detect these deceptive type of contents.

Transparency in the field of AI, however, raises significant debate, particularly regarding the disclosure of training data and algorithms. On one hand, transparency helps identify biases and helps the limitations of AI models used be understood, but it also poses risks to data security and confidentiality. Similarly, making algorithms accessible promotes accountability and combats discrimination, but at the same time, it could harm the competitiveness of businesses, especially at a time when these technologies represent a tremendous economic potential for European actors. Therefore, it is crucial to strike a balance between transparency and the protection of innovation to ensure the responsible use of artificial intelligence.

05.

Public Interest

MAKE.
ORG

AI ACTION
SUMMIT

Popular idea

Develop AI for health diagnostics

6 proposals

What the majority of citizens agree on

Creating AI dedicated to healthcare to improve disease diagnosis and risk assessment, providing effective support to doctors in analysing pathologies. This could also play a key role in combating serious diseases, such as cancer, not only in research but also in supporting treatments.

Developing tools to promote the autonomy and safety of elderly people at home is also essential. AI can optimise treatments, improve access to care, reduce costs, and detect epidemic trends.

❤️ Popular proposal examples:

Jeannine, 71 ans

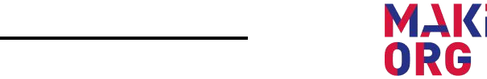
We should use AI as a legitimate assistant in fighting serious diseases (e.g., cancer), not just in research.

Jeannine, 71 ans

We should use AI as a legitimate assistant in fighting serious diseases (e.g., cancer), not just in research.

Fatima, 55 ans

We should develop tools that promote autonomy and security for elderly people living at home.



Angela, 55 ans

We should use AI to assist doctors in analysing pathologies.

Estelle, 56 ans

We should use AI to diagnose diseases, optimise treatments, improve healthcare access, and reduce costs.

Didier, 54 ans

We should use AI to improve medical diagnostics, manage epidemics, and identify data trends.

% votes "in favour" % “neutral” votes % votes "against"

Popular idea

Gradually use AI to optimise public services

5 proposals

What the majority of citizens agree on

Gradually integrating artificial intelligence into public services, while preserving the role of humans in understanding citizens' needs.

Using AI to simplify administrative tasks, make documents more accessible, and assist agents.

❤️ Popular proposal examples:

Jules

We should not dream of a grand AI breakthrough, but rather prioritize a gradual learning approach to improve our public services.

Benoit

We should ensure that AI is an efficient and secure tool to assist professionals in their administrative tasks.

Annie

We should ensure that public services are managed by humans in order to understand the needs of individuals.

Marie – Claire

We should automate the simplification of technical and administrative documents to make them accessible to the general public.

% votes "in favour" % “neutral” votes % votes "against"

Controversial idea

Address public interest issues through AI

17 proposals

What citizens are divided on

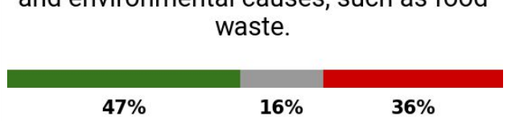
Mobilising AI to solve major global and social challenges - such as humanitarian crises, environmental crises, food shortages, animal protection, armed conflicts, and homelessness - through data modelling.

Creating tools and sharing them globally.

⚡ Controversial proposal examples:

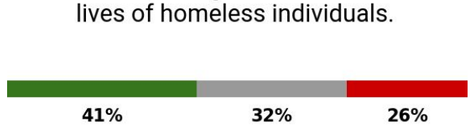
Philippe

We should develop AI within public institutions to address major social and environmental causes, such as food waste.



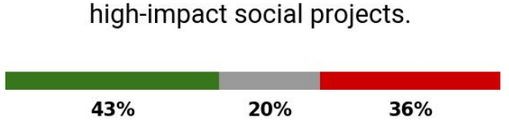
Théo, 24 ans

We should leverage AI to transform the lives of homeless individuals.



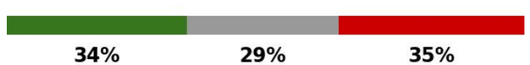
Camille, 36 ans

We should use AI to identify emerging sectors and direct investments toward high-impact social projects.



Bastien, 46 ans

We should develop AI solutions for animal protection.

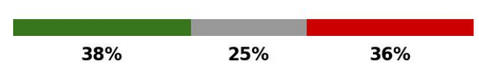


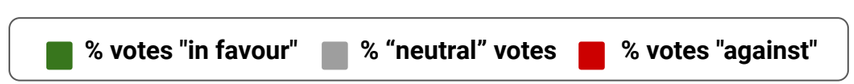
Lucas, 35 ans

We should develop AI to predict and prevent humanitarian crises and natural disasters.

Andre

We should leverage AI to prevent wars.





Controversial idea

Streamline public administration management with AI

14 proposals

What citizens are divided on

Automating tasks in public and administrative services to improve efficiency and, if necessary, streamlining positions, even replacing some.

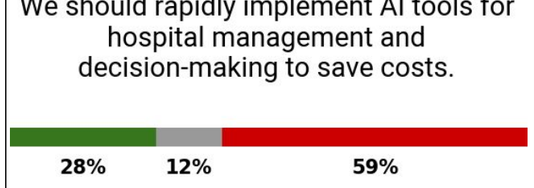
Fighting against bureaucracy using AI, including in the healthcare sector.

Optimising the management of local authority budgets and better coordinating the civil service.

⚡ Controversial proposal examples:

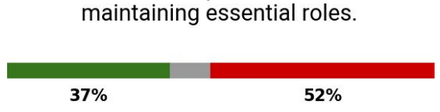
Paul, 64 ans

We should rapidly implement AI tools for hospital management and decision-making to save costs.



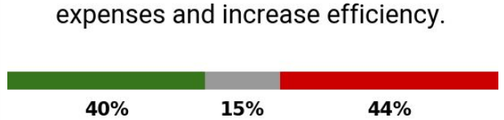
Jean, 40 ans

We should use AI to replace redundant civil servant positions while maintaining essential roles.



Jean, 40 ans

We should use AI to cut costs in certain administrative services to reduce expenses and increase efficiency.

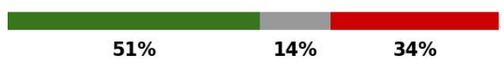


Rose, 60 ans

We should use AI to identify administrative inefficiencies and make the public sector more effective.

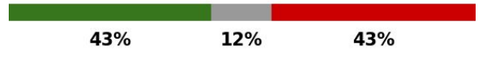
Sabrina

We should allow the public sector to use AI to reduce repetitive and time-consuming tasks.



Olivier, 28 ans

We should use AI to streamline administrative processes and regulations.



Controversial idea

Simplify access to justice and its functioning.

6 proposals

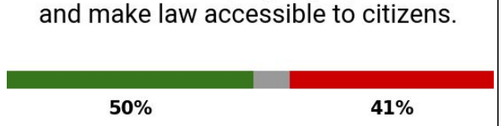
What citizens are divided on

Mobilising AI to simplify the judicial system by making the law more accessible through legal assistants that simplify procedures and chatbots that help people understand complex texts. AI could also accelerate investigations, including those on older cases, and assist judges in assigning penalties and managing repeat offenders, correcting inconsistencies and inefficiencies.

⚡ Controversial proposal examples:

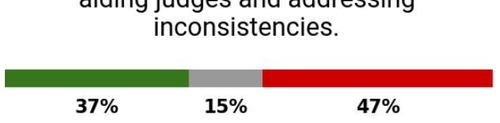
Estelle, 56 ans

We should create AI-based legal assistants to simplify legal processes and make law accessible to citizens.



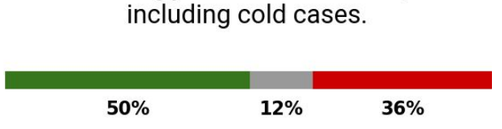
P, 62 ans

We should use AI to propose sentencing recommendations in criminal cases, aiding judges and addressing inconsistencies.



Jeannine, 71 ans

We should enable AI to play a major role in solving judicial cases faster, including cold cases.



Jean, 40 ans

We should use AI chatbots to help people understand complex legal, fiscal, and legislative texts.

Rose, 60 ans

We should use AI to assist judges in better managing and applying sentences for repeat offenders.

Controversial idea

Facilitate administrative procedures with the help of AI.

56 proposals

What citizens are divided on

Simplifying administrative procedures and everyday formalities.

Enabling more efficient tracking of requests and inquiries in real-time.

Providing aid simulators to quickly identify available support programmes.

Facilitating procedures for self-employed workers, from business creation to managing assistance.

⚡ Controversial proposal examples:

Jean, 40 ans

We should create AI tools to support new entrepreneurs with administrative tasks and documentation.

47% 20% 31%

Paul, 64 ans

We should create AI to assist users with all public service and administrative formalities.

52% 15% 32%

Olivier, 28 ans

We should use AI to streamline administrative processes and regulations.

43% 12% 43%

Paul, 64 ans

We should create AI to assist users with all public service and administrative formalities.

52% 15% 32%

Philippe

We should automate the categorisation and tracking of citizen requests using AI to enhance administrative responsiveness.

50% 14% 35%

% votes "in favour" % “neutral” votes % votes "against"

AI, a tool that serves public interest

"We must approach AI as a force – political, economic, cultural, and scientific... We must connect the issues of power and justice."
Kate Crawford, Professor at New York University, Researcher at Microsoft

Artificial intelligence is now seen not only as a technological tool, but also as a potential lever to solve complex societal problems and build a more equitable and sustainable society, effectively addressing citizens' needs and supporting public action. Its use holds untapped potential in many areas.

This is the case in healthcare, where AI can analyse large amounts of medical data to help diagnose diseases more quickly and with greater accuracy. It can also propose personalised treatment plans for each patient and accelerate research by identifying new drugs and analysing clinical trials more rapidly. In agriculture, AI can optimise the use of agricultural resources, improve crop yields, and reduce the environmental impact of farming through precision techniques. In transportation, AI can manage traffic in real time, thus reducing congestion and carbon emissions. AI-equipped autonomous vehicles could reduce road accidents by minimising human error. Regarding inclusion efforts, AI can develop assistive technologies for people with disabilities, such as screen readers for the visually impaired or smart prosthetics. By analysing socio-economic data, AI is also able to help identify disparities in access to resources and services.

In general, the development of predictive models allows social and environmental issues to be anticipated, the impact of large-scale decisions to be modelled, and helps our social and economic models anticipate and adapt to changes. In terms of security and justice, AI can analyse and cross-reference data to help law enforcement combat trafficking and detect fraud. AI is also capable of supporting the analysis of legal cases and recommending decisions based on laws or prior judgments. AI-powered drones and robots can be used to search for and rescue victims in dangerous or inaccessible areas. More broadly, AI can serve as a tool for decision-making and public action, particularly in the fight against fraud and trafficking.

Artificial intelligence could also bring significant improvements in quality, speed, and adaptability for public services. Its deployment in civil service administration helps reduce administrative tasks, allowing civil servants to focus more on their interactions with citizens. For citizens, AI can greatly simplify administrative procedures, make regulations and legal texts more accessible and understandable, personalise user support, translate messages addressed to citizens into any language, or automatically generate responses to questions asked online.

However, significant challenges remain, particularly the risk of reproducing or amplifying biases, which could lead to discrimination in sensitive areas such as justice, security (profiling, surveillance), or recruitment. Additionally, unequal access to technology exacerbates the digital divide, making services inaccessible to certain populations cut off from digital resources. Finally, the high costs of AI models and their monopolisation by a small number of actors risks increasing global imbalances, widening the gap between companies and countries. It is therefore crucial to adopt an ethical and responsible approach to ensure that artificial intelligence genuinely serves public interest, while respecting principles of fairness and inclusion.

06.

AI and the Environment

AI ACTION SUMMIT

Popular idea

Limit the environmental impact of AI.

10 proposals

What the majority of citizens agree on

Raising public awareness about the energy consumption required for the overall management of AI, such as cooling data centres.

Supporting the development of AI towards reduced impact: modernising data centres, launching a debate on benefits vs. impacts, and incorporating this into upcoming regulatory frameworks.

Halting the development of AI if it becomes too harmful.

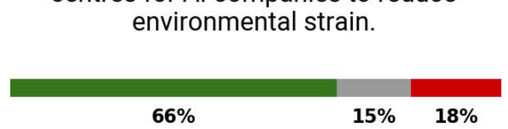
❤️ Popular proposal examples:

Sine, 58 ans

We should inform the public about the overheating caused by AI use: more cooling systems mean higher energy demand.

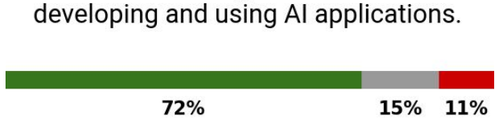
Anne, 45 ans

We should mandate carbon-neutral data centres for AI companies to reduce environmental strain.



Axa

We should ensure that environmental costs never outweigh the benefits when developing and using AI applications.



Sa, 50 ans

We should halt the development of ecologically damaging AI systems that foster human dependence.

Jean, 41 ans

We should embed environmental concerns into AI regulations and implementation strategies.

Marjorie

We should raise public awareness about the risks of AI and its environmental costs.

% votes "in favour" % “neutral” votes % votes "against"

Popular idea

Leverage AI to anticipate natural crises.

5 proposals

What the majority of citizens agree on

Using AI to detect and anticipate natural crises by utilising climate data through predictive models. It can also be integrated into the processes of organising emergency responses, managing populations, and coordinating interventions.

AI could also play a role in protecting biodiversity by monitoring ecosystems, identifying endangered species, and improving conservation strategies.

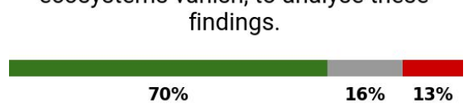
❤️ Popular proposal examples:

Marthe, 32 ans

We should leverage AI to analyse climate data and prevent natural disasters.

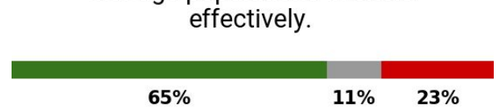
Francois, 50 ans

We should collect global wildlife data (e.g., cameras, observations) before ecosystems vanish, to analyse these findings.



Estelle, 56 ans

We should utilise AI to predict natural disasters, coordinate responses, and manage populations in crisis effectively.



Jeannine, 71 ans

We should harness AI to detect natural disasters well in advance to mitigate damage.

Marthe, 32 ans

We should leverage AI to analyse climate data and prevent natural disasters.

% votes "in favour" % “neutral” votes % votes "against"

Climate change: AI, cause or solution?

"Artificial intelligence (...) can accelerate sustainable development. Whether it's (...) helping farmers increase their yields, designing sustainable housing and transportation, or establishing an early warning system for natural disasters."
António Guterres, Secretary-General of the United Nations

AI amplifies the digital sector impact on the environment by causing notable environmental effects, primarily due to the increased energy consumption linked to its design (training of models) and usage: a query via a conversational agent like ChatGPT can require up to five times more energy than a traditional search. Energy demand could exceed that of air travel in a few years!

As a result, investments in low-carbon energy production technologies, such as miniature nuclear reactors, are being made, particularly by tech players. Additionally, the manufacturing of digital equipment accounts for about 60% of the ecological footprint of digital technology. The growing demand for computing hardware, especially to support applications like cloud computing platforms, data centres, and supercomputers, intensifies the pressure on natural resources. This raises concerns about the sustainability of rare metal supply and the management of electronic waste, which represents another major environmental challenge. Lastly, the exponential development of AI requires a more frugal approach, which questions real needs from the users' perspective, to be adopted.

But AI also offers opportunities to reduce environmental pressures. For example, Météo France uses models to predict weather conditions and anticipate natural disaster risks, such as flooding, heatwaves, and storms, thus improving emergency responses. AI is also used to monitor biodiversity by analysing satellite data to detect changes in ecosystems, such as deforestation in the Amazon, pollution, the loss of marine habitats, and coral reef degradation. These technologies enable a quicker and more targeted response to environmental crises while facilitating long-term planning for sustainable management of natural resources. In agriculture, AI also helps optimise resource usage by monitoring soil conditions and predicting yields, which reduces water and fertiliser waste.

Debates exist on the balance to be struck between environmental impact and the benefits of technology for adaptation and mitigation. Discussions on AI in relation to the environment often oscillate between its potential benefits and its ecological impacts. On one hand, applications like image generation for personal purposes are often seen as excessive energy consumption, which is often deemed unjustifiable in comparison to the environmental benefits. However, AI also offers significant opportunities for environmental protection.

Moreover, the limited existing regulation raises concerns about its use, particularly regarding transparency, performance, and the frugality of systems. The transparency of the databases on which AI is trained is crucial, particularly to confirm the absence of bias. This highlights the need for verification and regulation tools to ensure that AI technologies are used responsibly, particularly concerning their environmental impact and their ability to provide reliable information.

MAKE.

ORG

PART 2

Findings from the Expert Consultation

AI ACTION SUMMIT

Produced by The Future Society

THE FUTURE SOCIETY

The Expert Consultation Participants

- *Scientific Institutions*
- *Academic Centers*
- *Policy Think Tanks*
- *Labor Unions*
- *Artist Associations*
- *Local and global non-profits*

200+
experts and civil society organisations

5
continent represented

15
key deliverables

THE FUTURE SOCIETY

Methodology

Data Collection → **Interim Analysis** → **Final Analysis**

- Online survey with possibility to upload documents
- 200+ organizations invited
- Global outreach across 5 continents
- Promotion through Summit channels

- Review of first 84 submissions
- Identification of emerging themes
- Public sharing for feedback

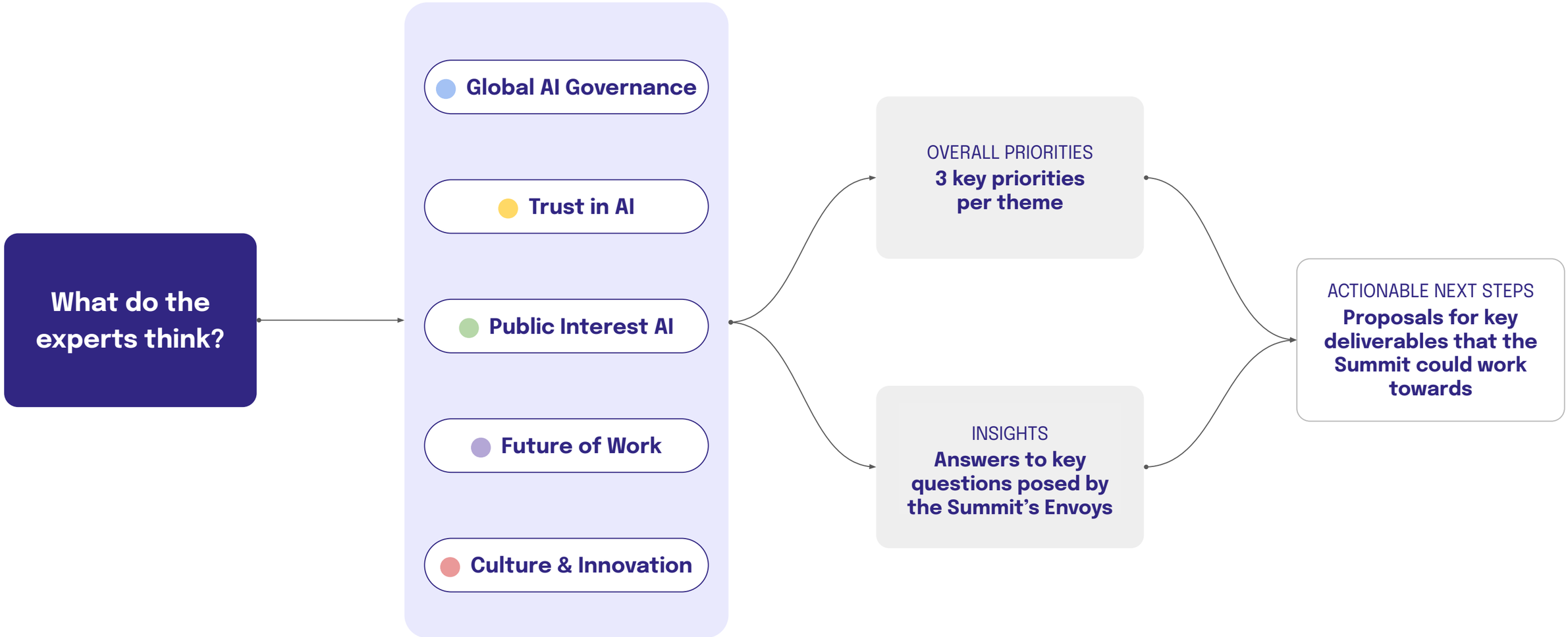
- 202 eligible submissions
- Analysis of priorities based on frequency of mentions
- Identification of recurring themes and patterns on key deliverables

15
Key Deliverables

20
Expert Insights

THE FUTURE SOCIETY

Structure of the Findings



THE FUTURE SOCIETY

Summary

THEME	EXPERT PRIORITY: WHAT SHOULD BE DONE?	DELIVERABLE: WHAT CAN THE SUMMIT DO?
Global AI Governance	Strategic Vision and Alignment for Future AI Summits	Assign a Permanent Governing Body for the Global AI Summits
	Increased Scientific Consensus on AI Risks, including on concepts of AI Transparency, Explainability, and Monitoring	Set the Foundation for the International Scientific Collaboration on AI
	Broader Inclusion in Global AI Governance	Design Mechanisms for Civil Society Organizations (CSOs) and Global South Inclusion
Trust in AI	Clear and Enforceable Red Lines for Advanced AI Development	Define and Publish Thresholds for AI Oversight
	Enhanced Responsible Scaling Policies that Build on AI Safety Commitments	Introduce the AI Corporation Commitments Report Card
	Stronger International Network of AI Safety Institutes	Develop a Collaborative Testing Framework for the AI Safety Institute Network
Public Interest AI	Unified Standards for Auditing AI Systems	Present a Roadmap for Global AI Auditing Standards
	Advancement of Public Interest AI Guidelines, including Responsible AI Practices	Draft and Adopt a Global Charter for Public Interest AI
	Clear Processes for Citizen Engagement in AI Governance	Launch the AI Commons Initiative to Empower Citizens in AI Design
Future of Work	Improved Global AI Literacy and Awareness of Risks and Opportunities	Create a Global AI Education and Critical Thinking Initiative
	Workforce Policies for AI Transition and Job Protection	Form an International Task Force on AI and Labor Market Disruption
	AI-Specific Educational and Training Programs	Establish a Global AI Talent Program for the Global South
Innovation & Culture	Enhanced Reporting and Promoting of Green AI Development	Announce the “GreenAI Leaderboard” to Track and Incentivize Model Sustainability Improvements
	Ethical Governance and Protection of Cultural Data Rights	Propose Standards for Cultural Data and AI Intellectual Property Rights
	Global Funding for AI Governance and Research	Coordinate a Fund for International Research Collaborations on AI with a Focus on Ethics, Alignment, and Equity

01.

Global AI Governance

98 expert and civil society organizations submissions

AI ACTION SUMMIT

THE FUTURE SOCIETY

Global AI Governance

EXPERTS' TOP PRIORITIES FOR THE THEME	EXAMPLES OF POTENTIAL KEY DELIVERABLES
1. **Strategic Vision and Alignment for Future AI Summits**	1. **Assign a Permanent Governing Body for the Global AI Summits**
2. **Increased Scientific Consensus on AI Risks, including on concepts of AI Transparency, Explainability, and Monitoring**	2. **Set the Foundation for the International Scientific Collaboration on AI**
3. **Broader Inclusion in Global AI Governance**	3. **Design Mechanisms for Civil Society Organizations (CSOs) and Global South Inclusion**

THE FUTURE SOCIETY

Deliverable

1. Assign a Permanent Governing Body for the Global AI Summits

ℹ️ Context

AI summits need clear follow-up. Global AI summits lack the structured mechanisms needed to transform high-level declarations into actionable, measurable strategies. This can lead to fragmented outcomes, limited capacity to monitor progress, and an absence of coordinated, long-term objectives for governing AI development–particularly for high-risk development and deployment.

Deliverable 🎯

- **Create a 15-member taskforce to guide future summits** and check results.
- **Give the taskforce the objective to establish high-level summit mechanisms:** multi-year roadmap, working groups, or structured coordination with global governance initiatives, such as those of the UN and the AI Safety Institutes Network
- Set clear process for **choosing summit hosts.**

Timeline 🗓️

Before the Summit

Pick taskforce members, write their rules, prepare announcement.

During the Summit

Launch the Paris Pact taskforce, start agreement work, share host selection rules.

After the Summit

First task force meeting in April, draft agreement, pick next host.

Expected benefits ⭐

🌍 **Global Leadership:** Establish a pioneering framework for coordinated AI governance

⚖️ **Clear Results:** Regular updates show which promises get kept

🤝 **Better Planning:** Each summit builds on previous work

📄 **Real Rules:** Track which commitments get kept

Footnotes

Inspired by submissions from: Safe AI Forum, EthicsNet, Missions Publiques, DTFTP, Future of Life Institute, Centre for International Governance Innovation, Connected by Data, Pour Demain, Open Future Foundation, Institute for AI Policy and Strategy, Ada Lovelace Institute, Oxford Martin AI Governance Initiative, Centre pour la Sécurité de l'IA (French Center for AI Safety), Ethical AI Alliance, Berggruen Institute, University of Burgundy, Observatoire de l'éthique publique, Institut Universitaire de France, Green IT, Handicap International - Humanity Inclusion, AGI Collective, Swedish Defence University, Intermobility, Renaissance Numérique, Center for a New American Security, Resilio, Existential Risk Observatory, Impact AI, Concordia AI, PauseAI, Federal, University of Montreal/Mila, Youth for Privacy

THE FUTURE SOCIETY

Deliverable

2. Set the Foundation for the International Scientific Collaboration on AI

ℹ️ Context

There is a lack of international scientific consensus on the risks posed by advanced AI systems and the appropriate governance mechanisms to mitigate those risks. Existing efforts, like the International Scientific Report on the Safety of Advanced AI, are a step toward that direction. However, there is a need for further development and institutional support to maximize its global legitimacy and impact.

Deliverable 🎯

- Establish the **International Scientific Collaboration on AI** to oversee the **State of Science AI Report**, ensuring balanced representation across governments, international bodies, industry, and academia.
- Formalize **UN partnership** through coordinated information sharing and aligned reporting mechanisms.
- Ensure comprehensive **stakeholder engagement**, particularly from **Global South** and **Asia-Pacific regions**.
- Develop **two-year strategic roadmap** with clear milestones and interim reporting requirements.

Timeline 🗓️

Before the Summit

Draft and circulate proposal for the International Scientific Collaboration structure and focus.

During the Summit

Convene stakeholders to formalize collaboration framework and determine institutional arrangements. Article the complementarity with existing mechanisms.

After the Summit

Launch implementation phase with designated working groups and coordination mechanisms.

Expected benefits ⭐

- 🔬 **Scientific Excellence:** Establish authoritative platform for rigorous analysis and assessment
- ⚖️ **Strategic Integration:** Align scientific insights with global policy frameworks through UN coordination
- 🤝 **International Consensus:** Foster enduring collaboration on AI science through structured engagement

Footnotes

Inspired by submissions from: Centre for International Governance Innovation, Oxford Martin AI Governance Initiative, Concordia AI, Centre pour la Sécurité de l'IA, Ethical AI Alliance, Ada Lovelace Institute, PauseAI, Active Service for the Benefit of Education and its Reform, Connected by Data, Impact AI, University of Montreal/Mila, PauseAI UK, Safe AI Forum, Global Partners Digital.

THE FUTURE SOCIETY

Deliverable

3. Design Mechanisms for Civil Society Organizations (CSOs) and Global South Inclusion

ℹ️ Context

Global AI summits and governance initiatives often lack meaningful participation from civil society organizations (CSOs) and representatives from the Global South, limiting the diversity of perspectives and hindering the development of equitable and inclusive AI.

Deliverable 🎯

- Design a plan for the **next Summit to be co-hosted by two countries**, a developing economy with the support of a developed economy.
- Create a dedicated fund to **support the participation of CSOs and Global South representatives** in AI Summits and related events.
- Enable **remote participation** to enable broader engagement from stakeholders who face logistical or financial barriers to attending in person.
- Launch a **permanent CSO advisory board** to provide input on AI governance and ensure that civil society perspectives are incorporated into decision-making processes.

Timeline 🗓️

Before the Summit

Consult with CSOs and Global South representatives on the advisory board, develop selection criteria for the advisory board, secure funding for participation support.

During the Summit

Announce the establishment of the advisory board, showcase best practices for inclusive participation, dedicate sessions for CSO and Global South perspectives.

After the Summit

Convene the advisory board, provide ongoing support for participation, integrate CSO and Global South input into AI governance initiatives.

Expected benefits ⭐

- ✨ **Increased diversity** of perspectives in AI governance.
- 🤝 More **equitable and inclusive** AI development and deployment.
- ⚖️ Enhanced **legitimacy and accountability** of AI governance initiatives.

Footnotes

Inspired by submissions from: Connected by Data, The Ada Lovelace Institute, Société française des traducteurs, Renaissance Numérique, Handicap International - Humanity Inclusion, Active Service for the Benefit of Education and its Reform, WITNESS, Digital Action, InternetLab, Fundação Itaú.

THE FUTURE SOCIETY

Insights

1. Current global AI governance ecosystem

🔎 Guiding questions

- What is the current state of global AI governance, including key actors, initiatives, and frameworks?
- What critical gaps exist in this ecosystem, and where do you see potential for synergy or unnecessary overlap?

☀️ Current Landscape

The global AI governance ecosystem is characterized by **fragmentation** across multiple actors and frameworks:

- **Intergovernmental organizations** (UN, OECD, UNESCO) primarily establish non-binding ethical guidelines and principles
- **Regional bodies** like the EU have implemented binding regulations through the AI Act, while the US has so far favored voluntary industry commitments
- **National governments** pursuing individual strategies with limited coordination
- **Tech companies** wielding significant influence through self-regulation and technical standards development

🚨 Critical Gaps

1. **Lack of enforceable global mechanisms**: Most frameworks remain **voluntary** or region-specific, limiting their effectiveness in addressing **global AI risks**.
2. **Underrepresentation of Global South perspectives**: This leads to risks of "**algorithmic imperialism**", where AI development primarily reflects Western interests and values.
3. **Insufficient focus on environmental impacts**: The energy consumption of large AI models is rarely accounted for.
4. **Inadequate preparation for catastrophic AI risks**: International governance approaches lacks urgency, despite expert warnings.

✨ Recommended Policy Actions

- **Harmonizing existing national AI strategies** and streamlining overlapping initiatives to enhance regulatory clarity and improve resource allocation efficiency
- **Strengthening cooperation** between major powers, especially the **US and China,** while ensuring meaningful inclusion of **Global South** nations in policy development
- **Establishing concrete enforcement mechanisms** including:
  - Mandatory reporting requirements
  - A centralized incident database for tracking AI-related issues
  - Shared enforcement protocols between nations
- **Incorporating environmental impact assessments into governance frameworks** and promoting sustainable AI development practices
- **Enhancing public engagement** through structured civil society participation in policy development and implementation

Footnotes

Inspired by submissions from: ANU School of Cybernetics, AI4DA, DTFTP, AI Voices, Future of Life Institute, Centre for ResponsibleAI IIT-Madras, Lincoln Alexander School of Law, Association Green IT, Resilio, Safe AI Forum, Existential Risk Observatory, Institute for AI Policy and Strategy, Connected by Data, Missions Publiques.

THE FUTURE SOCIETY

Insights

2. International Scientific Report on the Safety of Advanced AI

🔎 Guiding questions

- How can we best institutionalize and expand upon the International Scientific Report on the Safety of Advanced AI?
- What structures and processes are needed to ensure its ongoing relevance, objectivity, and impact on global policy decisions?

🤝 Consensus

The world needs a **permanent system to monitor and report on advanced AI development and deployment**, ensuring findings remain relevant and shape global policy effectively. Like climate change reporting, AI reporting requires a structured global approach with clear processes and expert oversight. The International Scientific Report on the Safety of Advanced AI is an important step toward driving scientific consensus, but it needs to be expanded and formalized.

📋 Proposed Governance Structure

- **Create a permanent governing body similar to the IPCC,** working under UN supervision for global coordination, as proposed by the UN Secretary-General, **or create another organization led by the AISI Network and focused on advanced AI.**
- Set up **regional committee**s for local insights and implementation. Build **advisory panels** of AI experts, ethics specialists, policy makers, and industry leaders
- Use clear, documented **peer-review processes** drawing on experts from multiple fields. Make all data sources and research methods public for **independent verification**

🗂️ Operational Framework & Resources

Key Elements:

- **Update findings yearly** to capture new AI developments concerns, ideally with interim reports every 6 months
- **Build a worldwide database** to share AI research findings
- **Run a live monitoring lab** to track data security and system performance
- **Publish all materials in multiple languages** for global access

Funding Plan:

- **Stay independent** by securing money from governments, international organizations, and businesses partnerships
- **Set up clear rules for how funding is managed and used**

✒️ Policy Impact & Integration

- Write **specific policy recommendations** that connect to global rules and potential AI treaties
- Meet regularly with policy makers
- **Work with existing AI oversight groups** to avoid duplicate efforts;
- Help shape **international AI standards**
- Get regular feedback from public organizations, businesses, and researchers

Footnotes: Inspired by submissions from: EthicsNet, DTFTP, Future of Life Institute, AI Safety Asia, Centre for ResponsibleAI (IIT-Madras), Active Service for the Benefit of Education and its Reform, Centre for International Governance Innovation, Connected by Data, Pour Demain, Institute for AI Policy and Strategy, Ada Lovelace Institute, OpenMined Foundation, Oxford Martin AI Governance Initiative, Centre pour la Sécurité de l'IA (French Center for AI Safety), Ethical AI Alliance, Stanford, Berggruen Institute, University of Burgundy, Observatoire de l'éthique publique, Institut Universitaire de France, Youth for Privacy, Association Green IT, Handicap International - Humanity Inclusion, Carnegie Endowment for International Peace, Alexander von Humboldt Institut für Internet und Gesellschaft, Swedish Defence University, Rick Gillespie, InterAgency Institute, Lusófona University, Mykolas Romeris University, Global Partners Digital, ANU School of Cybernetics, CIFAR, Individual, University of Montreal/Mila, Center for a New American Security, Existential Risk Observatory, Impact AI, Concordia AI, PauseAI, Open Markets Institute, ITU-T Focus Group on AI-Native Networks (FG AI-Native), University of Oxford, Reporters Without Borders (RSF), University of Cambridge, Legal Priorities Project, AI Governance Institute, France AI Commission.

THE FUTURE SOCIETY

Insights

3. Global body for AI oversight

🔎 Guiding questions

- What are the potential benefits, challenges, and key considerations in establishing a global body for AI oversight?
- How should such a body be structured, governed, and empowered to effectively address international AI challenges?

☝️ What would a global body for AI oversight do?

This global body would **set clear standards for AI alignment, ethics, and governance** while promoting international cooperation. This organization would verify compliance with ethics and alignment measures and protect against risks like AI misuse and weaponization.

🏛️ Recommended Structure

1. **Leadership and Organization**
   - Include all key groups: governments, public, and experts
   - Main council with changing leadership to ensure fairness
   - Open decision-making to build trust
2. **Daily Operations**
   - Clear rules with real consequences for violations
   - Technical experts to check compliance
   - Regular input from all affected groups
   - Power to conduct international inspections
   - Regular updates to rules as AI develops

🪜 Major Challenges

- **Finding agreement between countries** with different interests and concerns about control
- Preventing excessive bureaucracy from **slowing down progress**
- **Protecting against control** by powerful countries or companies
- Building **worldwide agreement** on rules and enforcement
- Keeping up with **rapid AI advances** while maintaining oversight

⭐ Key Benefits

- Single global framework to **set and enforce AI safety standards**
- **Strong verification system** for safety compliance, built on IAEA model
- Better **handling of AI risks** that cross national borders
- Unified rules to **maximize benefits** and r**educe risks** through clear standards

Footnotes:

Inspired by submissions from: Safe AI Forum, EthicsNet, Future of Life Institute, Centre for International Governance Innovation, Swedish Defence University, Renaissance Numérique, Center for a New American Security, Existential Risk Observatory, PauseAI

THE FUTURE SOCIETY

Insights

4. AI Summits

🔎 Guiding questions

- How can we enhance the effectiveness and impact of global AI summits?
- What specific measures would ensure these events drive meaningful progress in AI governance and foster sustained international cooperation? Which countries could organize the next ones?

⚛️ Essential Elements for Success

- **Create specific, time-bound commitments** backed by permanent working groups on AI alignment, ethics, and governance to drive continuous progress
- **Ensure broad representation through direct involvement** of Global South nations, governments, tech industry, researchers, and civil society organizations
- **Build public trust** through open reporting and community input processes
- **Develop two-way partnerships** between nations to strengthen worldwide AI collaboration

🔝 Key Strategic Priorities

- Secure ongoing **Chinese participation** for effective global governance
- Implement **joint hosting** between developed and developing nations to share expertise
- **Move summit locations between regions** to capture diverse viewpoints
- Include **countries with varied political ties** to strengthen international dialogue
- Prioritize Global South input to **build truly inclusive AI frameworks**

🗺️ Primary Host Country Candidates

- **Singapore**: Stands out for three key strengths: politically neutral stance, cutting-edge AI capabilities, and unique position to connect Eastern and Western perspectives. Location makes it ideal for building Asian cooperation networks
- **Canada**: Distinguished by world-class AI research and proven track record with Global Partnership on AI. Strong infrastructure through institutions like MILA and clear commitment to responsible AI practices
- **India**: Rapidly expanding AI sector with focus on solving social challenges
- **Kenya**: Emerging as Africa's AI policy leader with strong focus on ethical development as a member of the AISI Network
- **Brazil**: Leading Latin American AI development with growing technology base
- **Switzerland**: Distinguished by long-standing neutrality, strong research institutions, and proven track record hosting international organizations. Scientific diplomacy expertise and stakeholder facilitation capabilities make it particularly suitable
- **UAE**: Major AI investor with dedicated government focus through AI Ministry

Footnotes:

Inspired by submissions from: Safe AI Forum, EthicsNet, AI Safety Asia, Ada Lovelace Institute, Centre for International Governance Innovation, DTFTP, Future of Life Institute, Pour Demain, Institute for AI Policy and Strategy, Centre for ResponsibleAI, Impact AI, Ethical AI Alliance, PauseAI, Center for a New American Security, Berggruen Institute

THE FUTURE SOCIETY

Insights

5. Global approach with all countries

🪜 The Challenge

The international community faces a critical task: making sure every country, regardless of their current AI capabilities, **can help shape AI governance and share in its benefits**. This requires coordinated action on education, governance, and resource sharing.

🧰 Key Strategic Pillars

🔎 Guiding questions

- How do we ensure that all countries can have a say in the governance of AI and enjoy the technologies' benefits, including those with limited access to AI resources?

1. Building Local AI Expertise 💡

- **Create AI technology centers** across Africa, Latin America, and Southeast Asia to build regional decision-making power
- **Launch practical training programs** to strengthen AI knowledge and skills at all levels
- **Foster partnerships** between universities and businesses to develop local talent

2. Fair Global Governance ⚖️

- Improve the current AI international governance system to **give a fair voice** to every country.
- Share leadership roles among different regions to **prevent any group from dominating**
- Establish targeted funding streams to **help developing nations participate** fully

3. Sharing Resources Worldwide 🌍

- **Launch a Global AI Fund**, supported by advanced economies and tech companies, to assist developing regions
- **Build a shared platform** offering free access to AI tools and data, with clear safety guidelines
- Create **partnerships between government and industry** to share technology and expertise

Footnotes:

Inspired by submissions from: DTFTP, Future of Life Institute, Ada Lovelace Institute, AI Voices, Ethical AI Alliance, Swedish Defence University, ANU School of Cybernetics, Pulse GROUPE SOS, Impact AI, Concordia AI, AiXist, Audere, University of Oxford, Centre for International Governance Innovation, Centre for ResponsibleAI (IIT-Madras), Renaissance Numérique, Active Service for the Benefit of Education and its Reform, InterAgency Institute and Lusófona University

02.

Trust in AI

88 expert and civil society organizations submissions

THE FUTURE SOCIETY

Trust in AI

EXPERTS' TOP PRIORITIES FOR THE THEME	EXAMPLES OF POTENTIAL KEY DELIVERABLES
1. **Clear and Enforceable Red Lines for Advanced AI Development**	1. **Define and Publish Thresholds for AI Oversight**
2. **Enhanced Responsible Scaling Policies that Build on AI Safety Commitments**	2. **Introduce the AI Corporation Commitments Report Card**
3. **Stronger International Network of AI Safety Institutes**	3. **Develop a Collaborative Testing Framework for the AI Safety Institute Network**

THE FUTURE SOCIETY

Deliverable

1. Define and Publish Thresholds for AI Oversight

ℹ️ Context

Following the Seoul AI Summit's mandate, there is a strong expert consensus on developing clear, actionable definitions of thresholds triggering oversight as well as intolerable risk thresholds for AI systems, balancing innovation with essential safety guardrails.

Deliverable 🎯

- Define **practical risk thresholds** with experts from industry, academia, and civil society
- Develop **assessment framework focused on measurable metrics** such as compute and specific dangerous capabilities such as autonomous replication, self-improvement, or advanced manipulation
- Create implementation **guidelines based on existing regulatory frameworks** in finance and healthcare
- Establish voluntary **monitoring system** for high-capability AI systems
- Produce **concise report** with clear thresholds and industry-ready monitoring approaches

Timeline 🗓️

Before the Summit

Draft initial intolerable risk thresholds through consultation with experts, the Network of AI Safety Institutes and existing frameworks such as the EU AI Act, building on existing work such as the OECD's public consultation

During the Summit

Present draft, gather feedback from key stakeholders, refine framework

After the Summit

Begin voluntary implementation with major AI companies, develop implementation roadmap, explore early warning systems to identify concerning AI capabilities before deployment.

Expected benefits ⭐

- ✨ Clear **expert consensus** on unacceptable AI risks
- ⚙️ Practical mechanisms for **identifying high-impact systems**
- 🤝 Enhanced **coordination** between developers and regulators
- ⚖️ Balance between **innovation and responsible development**

Footnotes

Inspired by submissions from Rural Empowerment and Institutional Development (REPID), SKEMA Business School, EthicsNet, DTFTP, KETI, Alter.org.il, Future of Life Institute, SaferAI, University of Oxford, Pour Demain, Institute for AI Policy and Strategy, Oxford Martin AI Governance Initiative, Centre pour la Sécurité de l'IA, Convergence Analysis, RSA Conference, Impact AI, International Center for Future Generations, Concordia AI, Learning Lab for Resiliency, TRAIL (Trusted AI Labs), Digital Action, Existential Risk Observatory, DFKI, and the Center for Long-Term Cybersecurity (CLTC) at UC Berkeley, Center for a New American Security, Johns Hopkins Center for Health Security.

THE FUTURE SOCIETY

Deliverable

2. Introduce the AI Corporation Commitments Report Card

ℹ️ Context

AI companies have made numerous global commitments for trustworthy AI at previous summits, notably in Seoul. However, there aren't any accountability mechanisms to ensure these commitments are being adhered to.

Deliverable 🎯

- A practical **assessment framework** focused on concrete, achievable metrics
- Individual company reports detailing their progress on each commitment.
- Industry best practices and successful advancements in **AI safety**

📁 Existing frameworks to build on:

EU Code of Practice; UNESCO Recommendation on the Ethics of AI

Timeline 🗓️

During the Summit

Host a dedicated session for companies to present their progress reports and engage in discussions with stakeholders. Publish the final report, including recommendations for future action.

After the Summit

Begin voluntary implementation with major AI companies, develop implementation roadmap, explore early warning systems to identify concerning AI capabilities before deployment.

Expected benefits ⭐

- ⏫ Increased transparency and accountability in the **AI industry**.
- 💡 Identification of best practices and areas for improvement.
- 🤝 Strengthened **trust** between AI companies and the public.

Footnotes

Deliverable inspired by the submissions of: Rural Empowerment and Institutional Development (REPID), DTFTP, KETI, Alter.org.il, Future of Life Institute, SaferAI, Impact AI, CLTR, Confiance.ai, OpenMined, Oxford Martin School, Centre pour la Sécurité de l'IA, Concordia AI, TRAIL (Trusted AI Labs).

THE FUTURE SOCIETY

Deliverable

3. Develop a Collaborative Testing Framework for the AI Safety Institute Network

ℹ️ Context

Following the November 2024 meeting in San Francisco, the International Network of AI Safety Institutes has established priority areas in driving scientific consensus, and evaluations and testing of frontier models. There is now a need to operationalize these commitments into concrete testing frameworks and protocols. To do so effectively, there is also a need for the AISI Network to establish a robust coordination arm that allows for knowledge exchange and joint testing collaboration.

Deliverable 🎯

- Design and expand shared protocols for joint testing exercises between member AISIs
- Create standardized templates and methodologies for sharing domestic evaluation insights
- Develop guidance framework for consistent interpretation of test results
- Build technical documentation in multiple languages to enable broader participation

Previous Work ↩️

- INASI November 2024 first meeting outcomes
- Seoul Statement of Intent framework
- Existing national AI safety evaluation approaches

Timeline 🗓️

Before the Summit

- Compile existing testing approaches from member institutes
- Build on the existing pilot testing exercise from the International Network of AI Safety Institutes
- Prepare multilingual documentation template

During the Summit

- Agree on scope of the first round of joint testing exercises
- Establish working groups for implementation

After the Summit

- Begin the collaborative testing efforts planned during the Summit
- Implement feedback mechanisms
- Scale successful approaches

Expected benefits ⭐

- ⚙️ Practical implementation of the testing collaboration goals of the AISI Network
- ⏫ Increased participation from regions at different stages of AI development
- 💡 More consistent interpretation of safety evaluation results across member institutes

Footnotes

Inspired by submissions from: DTFTP, KETI, Convergence Analysis, Oxford Martin AI Initiative, French Center for AI Safety, Oxford Martin AI Governance Initiative, Centre pour la Sécurité de l'IA (French Center for AI Safety), Pour Demain, OpenMined Foundation, International Center for Future Generations, DFKI, METR, Johns Hopkins Center for Health Security, SaferAI, Concordia AI, ETHIQAIS, International Dialogues on AI Safety (IDAIS), Centre for Long Term Resilience (CLTR)

THE FUTURE SOCIETY

Insights

1. Understanding of AI

🔎 Guiding question

- How can we improve the scientific understanding of AI risks and opportunities?

📌 Premise

To improve the scientific understanding of AI risks and opportunities, a multi-pronged approach focusing on interdisciplinary research, transparency, and international collaboration is essential.

Other Notable Suggestions:

- Investing in **education programs about AI ethics** and alignment to prepare future researchers.
- Implementing ongoing assessment frameworks to **monitor AI systems' performance and impact**, adapting policies as new risks and opportunities emerge.
- Establishing clear security and ethical thresholds to regulate the development and deployment of **AI systems**.
- Conducting long-term studies to observe the **effects of AI** over time, which can reveal how AI affects society over time.
- Engaging the **public in discussions about AI** to ensure that societal values and concerns are considered in AI development and regulation.

1. Fostering interdisciplinary research 🔎

This means encouraging collaboration between AI experts and professionals from various disciplines, such as ethics, sociology, psychology, and law. This collaborative approach can offer a comprehensive view of AI's influence on society, going beyond purely technological considerations. It should also build on the International Scientific Report on the Safety of Advanced AI, as discussed in the Global AI Governance section of this report.

2. Transparency in AI development and research 🛡️

Encouraging open access to research, datasets, and models can enable broader scrutiny of AI systems, thereby helping researchers identify potential risks and explore opportunities for innovation. Open data-sharing frameworks can facilitate this transparency and allow for validation and expansion of existing work.

3. Enhancing international collaboration to ensure a globally harmonized approach to AI governance 🤝

Establishing a formal structure for the International Network of AI Safety Institutes can facilitate the sharing of knowledge and best practices, enabling coordinated efforts in evaluating AI risks and opportunities.

Footnotes:

Inspired by submissions from: DTFTP, EthicsNet, KETI, SaferAI, Oxford Martin AI Governance Initiative, Centre pour la Sécurité de l'IA (French Center for AI Safety), Convergence Analysis, CIFAR, TRAIL (Trusted AI Labs), Pour Demain, Data & Society, Centre for Long Term Resilience (CLTR), Confiance.ai, Minderoo Centre for Technology and Democracy, Learning Lab for Resiliency®, University of Oxford, Australian National University School of Cybernetics, Impact AI, International Center for Future Generations, Concordia AI, Rural Empowerment and Institutional Development (REPID), SKEMA Business School, I L Expansions

THE FUTURE SOCIETY

Insights

2. Standards on AI

🔎 Guiding questions

- What specific technical standards and benchmarks for AI trustworthiness, explainability, and fairness should be developed and implemented globally?
- How can we ensure these standards are adaptable to rapid AI advancements?

🌐 Technical Standards & Adaptability

• **Monitor and Update:** review standards regularly to keep them current.

• **Flexible Framework:** Design rules that evolve with the state of the art, such as the EU AI Act.

• **Cross-sector Coordination:** Foster company, CSOs and government collaboration

• **Research Integration:** Use latest findings to enhance AI standards.

🔐 Trustworthiness

• **Security and Privacy**: Establish robust protocols for data protection, encryption, and secure data handling practices to ensure AI systems are resilient against cyber threats and prioritize user privacy.

• **Reliability and Robustness**: Develop benchmarks for the reliability and robustness of AI systems, including stress testing and failure mode analysis to guarantee consistent and accurate performance across diverse scenarios.

• **Transparency**: Implement guidelines for transparent AI development processes, encompassing documentation of design choices, data sources, and model performance to enable audits and accountability.

🗣️ Explainability

• **Model Interpretability**: Develop standards for developing interpretable AI models capable of providing clear and understandable explanations for their decisions, especially in high-stakes domains like healthcare and finance.

• **User-Friendly Explanations**: Establish benchmarks for generating explanations tailored to different user types, ensuring that both technical experts and non-experts can comprehend AI decision-making processes.

• **Auditability**: Develop guidelines for creating comprehensive audit trails that facilitate the examination and verification of AI decision-making, contributing to greater transparency and accountability.

⚖️ Fairness

• **Bias Detection and Mitigation**: Implement standards for detecting and mitigating harmful biases in AI systems, encompassing fairness metrics and bias correction techniques to ensure equitable outcomes for all individuals and groups.

• **Inclusive Data Practices**: Set benchmarks for ensuring diverse and representative datasets used in training AI models, minimizing biases stemming from unrepresentative data and promoting fairness across various demographics.

• **Equity Impact Assessments**: Implement guidelines for conducting equity impact assessments to evaluate potential disproportionate impacts of AI systems on different demographic groups, mitigating potential harms.

THE FUTURE SOCIETY

Insights

3. Thresholds

🔎 Guiding questions

- How can we define and implement clear, enforceable thresholds for AI development and deployment, considering both current and potential future capabilities?
- What mechanisms should be in place to regularly review and update these boundaries?

📖 Defining and Implementing Enforceable Thresholds for AI Development and Deployment

To ensure the responsible development and deployment of AI, **a multi-tiered framework for setting clear, enforceable thresholds is critical**. This framework should consider the system capabilities, compute resources, risks, and societal impact of AI systems, taking into account both current and potential future advancements.

📊 Multi-Tiered Threshold Framework:

- **Training Compute Power**: Using quantifiable metrics like training compute power as an initial filter can help identify systems needing closer oversight. As compute power is a significant driver of AI advancements, setting thresholds based on this metric can provide an early indication of potentially high-risk AI systems.
- **Capability-Based Evaluations**: Once identified, systems can be further evaluated based on their capabilities, focusing on potentially dangerous abilities such as autonomous replication, expert-level manipulation, or the ability to facilitate bio-synthesis. These evaluations should encompass both a model's demonstrated capabilities and its propensities, accounting for its likelihood of engaging in certain behaviors.
- **Societal Risk-Based Thresholds**: Broader societal risks should also inform threshold-setting. These thresholds should be established based on societal risk tolerances, informed by industry standards and expert consensus, considering potential harm to human safety, rights, and well-being.

⚙️ Mechanisms for Review and Update:

- **Regular Reviews**: To ensure these thresholds remain relevant as AI capabilities evolve, regular reviews involving diverse stakeholders are essential. This would involve:
  - Periodic audits of AI systems to assess compliance with standards.
  - Engaging AI developers, users, policymakers, ethicists, and the public in the review process.
- **State of the Art Regulations**: Regulatory frameworks need to be adaptable and dynamic to reflect advancements in AI technology and emerging risks - the EU AI Act being an example of an updatable regulation..
- **International Coordination**: International coordination is crucial to harmonize standards, facilitate information exchange, and align risk assessment methodologies globally, including through collaborating with the OECD as a continuation of the OECD's effort on risk thresholds. This can help reduce compliance costs for AI developers while fostering international collaboration on AI safety and governance.

Footnotes:

Inspired by submissions from: Future of Life Institute, SaferAI, University of Oxford, Pour Demain, Oxford Martin AI Governance Initiative, Centre pour la Sécurité de l'IA (French Center for AI Safety), Concordia AI, Impact AI, International Center for Future Generations, Center for Long-Term Cybersecurity (CLTC) at the University of California, Berkeley, Convergence Analysis, Center for a New American Security, Existential Risk Observatory, Rural Empowerment and Institutional Development (REPID), Digital Action.

THE FUTURE SOCIETY

Insights

4. Follow-up on past commitments

🔎 Guiding question

- What concrete mechanisms and reporting standards should be established to ensure AI companies transparently demonstrate adherence to global AI ethics commitments they made at previous Summits?

🕵️ Third-Party Audits & Transparency Reports:

AI companies should undergo regular third-party audits to verify compliance with ethical guidelines on fairness, accountability, and bias mitigation. Audit results should be published in **annual transparency reports**, promoting accountability and allowing for public scrutiny.

🗃️ Centralized Public Registry:

AI Impact Assessments, evaluating risks like bias and privacy, should be submitted to a centralized public registry. This registry would enable **real-time monitoring of ethical adherence** and provide valuable insights for policymakers and researchers.

🌍 International AI Ethics Board:

An international AI ethics board could be established to oversee the submitted reports, providing independent assessments and facilitating global collaboration on ethical AI development and deployment. This board could also play a crucial role in **updating ethical standards as AI technology advances**.

📊 Public-Facing Dashboards:

Public-facing dashboards tracking **AI systems' societal impacts** and **ethical incidents** would enhance transparency and allow for public engagement in AI governance. These dashboards could also incorporate early-warning systems to alert regulators when models approach risk thresholds, enabling proactive intervention and mitigation strategies.

👩‍💻 Independent Observatory:

Establishing an independent observatory to track **global progress** and **periodically evaluate ethical standards** would ensure that global AI ethics commitments evolve alongside technological advancements. This observatory could collaborate with an international AI ethics board to provide recommendations for policy adjustments and updates to guidelines.

Footnotes:

Inspired by submissions from: SKEMA Business School, EthicsNet, DTFTP, KETI, Future of Life Institute, Oxford Martin AI Governance Initiative, Centre pour la Sécurité de l'IA (French Center for AI Safety), Johns Hopkins Center for Health Security, Centre for Long Term Resilience (CLTR), Confiance.ai, Impact AI, Concordia AI, CIFAR, University of São Paulo.

THE FUTURE SOCIETY

Insights

5. Cooperation and international network of AI Safety Institutes

🔎 Guiding questions

- How can we foster effective international cooperation in AI governance, including the creation of a global network of AI evaluation and safety institutes?
- What should be the key objectives, structure, and funding mechanisms for such a network?

The International Network of AI Safety Institutes (AISIs) was officially launched on November 20, 2024, after this consultation concluded. While the network has already begun joint testing exercises, the experts' insights remain valuable for its continued evolution and expansion.

🔑 Strategic Priorities:

- **Developing Standardized Evaluation Criteria:** Building on current joint testing initiatives, the expanded network should continue developing standardized evaluation criteria for AI safety, ethics, and risk assessment. This includes establishing clear, measurable thresholds for evaluating potentially dangerous AI capabilities.
- **Promoting Transparency and Accountability:** The network should strengthen transparency and accountability in AI development by facilitating the sharing of best practices, methodologies, and risk assessment frameworks among member institutes.
- **Capacity Building:** A crucial priority remains building capacity by providing training, resources, and technical expertise to countries, particularly those in the Global South, to develop their own AI governance capabilities.

🖇️ Structure of the Network:

- **Building on Existing Foundations:** The International Network of AISIs represents an important step toward coordinated scientific consensus building. This foundation can be expanded by establishing clear membership criteria and actively working to ensure geographical and institutional diversity.
- **Complementary Structures:** A global consortium modeled after the U.S. AISI consortium could complement the existing network by including laboratories, civil society organizations, and academic institutions. This broader ecosystem could be affiliated with established international organizations like the UN or OECD to enhance legitimacy and coordination.

💰 Funding Mechanisms:

A sustainable funding model is crucial for the long-term success of the network. Funding could be secured through a combination of government contributions, private sector partnerships, philanthropic grants, and membership fees.

Footnotes: Inspired by submissions from: Rural Empowerment and Institutional Development (REPID), DTFTP, KETI, SaferAI, Oxford Martin AI Governance Initiative, Centre pour la Sécurité de l'IA (French Center for AI Safety), Johns Hopkins Center for Health Security, Convergence Analysis, Centre for Long Term Resilience (CLTR), Confiance.ai, Impact AI, SKEMA Business School, EthicsNet, Safe AI Forum, AI 4 Development Agency (AI4DA), University of Toronto and Vector Institute, Reporters Without Borders (RSF).

03.

Public Interest AI

79 expert and civil society organizations submissions

THE FUTURE SOCIETY

Public Interest AI

EXPERTS' TOP PRIORITIES FOR THE THEME	EXAMPLES OF POTENTIAL KEY DELIVERABLES
1. **Unified Standards for Auditing AI Systems**	1. **Present a Roadmap for Global AI Auditing Standards**
2. **Advancement of Public Interest AI Guidelines, including Responsible AI Practices**	2. **Draft and Adopt a Global Charter for Public Interest AI**
3. **Clear Processes for Citizen Engagement in AI Governance**	3. **Launch the AI Commons Initiative to Empower Citizens in AI Design**

THE FUTURE SOCIETY

Deliverable

1. Present a Roadmap for Global AI Auditing Standards

ℹ️ Context

The rapid deployment of AI systems has outpaced the development of standardized, transparent, and independent auditing mechanisms, leaving significant gaps in accountability and public trust. Existing practices lack comprehensive oversight across the AI lifecycle, increasing risks and undermining ethical compliance globally.

Deliverable 🎯

- Develop and implement international standards for **AI auditing and monitoring**. This includes pre-deployment risk evaluations, continuous monitoring of live **AI systems**, independent audits (including red-teaming exercises), safety reporting protocols, and standardized post-market impact assessments. Create governance mechanisms covering the entire AI lifecycle.
- Establish an audit framework for AI companies. This framework should include:
  - A standardized reporting template based on **UNESCO** and **EU** guidelines
  - Guidance on verifying AI systems and processes, covering data usage, algorithmic fairness, and compliance with **ethical guidelines**.
  - Clear **KPIs** for measuring adherence to ethical principles
  - Independent verification mechanisms and regular audits
  - A **public-facing platform** tracking implementation progress

Timeline 🗓️

Before the Summit

Draft audit framework and reporting templates

During the Summit

Finalize standards through stakeholder consultation

After the Summit

Launch public tracking platform and begin first audit cycle; Establish an international verification system

Expected benefits ⭐

- 🔄 Continuous **monitoring of AI ethics** implementation
- ⚙️ Standardized **industry-wide** reporting practices
- 🔍 Independent **verification** of company claims
- 🌟 Enhanced **public oversight** of AI development

Footnotes

Inspired by: Rural Empowerment and Institutional Development (REPID), DTFTP, KETI, Future of Life Institute, SaferAI, Impact AI, CLTR, Confiance.ai, OpenMined, Oxford Martin, Centre pour la Sécurité de l'IA, Concordia AI, TRAIL (Trusted AI Labs), Safe AI Forum, Centre for Responsible AI (IIT-Madras), Connected by Data, Carnegie Endowment for International Peace, Laboratoire de l'Égalité, DFKI, Digital Scales, Z-inspection Initiative, Center for a New American Security, AiXist, InternetLab, Shift Project, University of Oxford, Center for Long-Term Cybersecurity

THE FUTURE SOCIETY

Deliverable

2. Draft and Adopt a Global Charter for Public Interest AI (1/2)

ℹ️ Context

A lack of a unified definition of “public interest AI” hinders the development and deployment of beneficial AI systems. Without a definition, several challenges persist, including inconsistent policy and regulation, difficulty in prioritizing research and a lack of clarity for developers in determining how to align their systems with the public interest.

Deliverable 🎯

- Draft a **Public Interest AI Charter** that defines “public interest AI” and outlines shared values, goals, and outcomes.
- Ensure the **inclusion of impacts** on a variety of groups, including children, indigenous communities, youth, women and people of all genders, precarious workers, and any other marginalized group.
- Include specific, measurable metrics to evaluate **AI systems**' adherence to the charter.
- Create a comprehensive **digital charter platform** featuring:
  - Self-assessment toolkit for organizations
  - Public commitment tracker for signatories
  - Real-time dashboard of adoption metrics

Timeline 🗓️

Before the Summit

Convene a working group to draft the charter.

During the Summit

Present the draft charter for discussion and feedback.

After the Summit

Finalize and publish the charter.

Previous Work & Stakeholders to Involve ↩️

Existing definitions of “**public interest AI**,” frameworks like the UN Sustainable Development Goals, organizations like UNESCO, the OECD, and the EU, and experts in AI ethics and governance.

Expected benefits ⭐

- 🤝 A shared understanding of “public interest AI” among stakeholders.
- 💡 Guidance for the development and deployment of AI systems that serve the public good.

Footnotes

Inspired by submissions from: Rural Empowerment and Institutional Development (REPID), Future of Life Institute, SaferAI, Centre for Responsible AI (IIT-Madras), Connected by Data, Global Forum for Media Development (GFMD), University of Oxford, Open Data Charter, InternetLab, Saskatchewan Indigenous Cultural Centre, 5Rights Foundation, Handicap International - Humanity Inclusion, Safe AI For Children Alliance, HEAT project, Züger et al

THE FUTURE SOCIETY

Deliverable

2. Draft and Adopt a Global Charter for Public Interest AI (2/2)

Example of a Public Interest AI Charter

💡 Definition

➢ Public Interest AI systems measurably advance societal well-being while protecting individual rights and cultural values, demonstrated through verifiable mechanisms and transparent assessment.

📝 Review Process

- Annual multi-stakeholder review
- Quarterly metric assessment
- Continuous public feedback

📐 Three Pillars & Metrics

1. **Societal Benefit**
   - Alignment with UN SDGs
   - Measurable public good outcomes
   - Public sector effectiveness scores
2. **Rights Protection**
   - UNESCO AI Ethics compliance
   - Bias & fairness indicators
   - Accessibility rates
3. **Implementation Standards**
   - Impact assessments
   - Cultural adaptability to diverse contexts while maintaining consistent ethical standards
   - Clear redress mechanisms

👥 Public Interest AI Examples

➢ Disease screening for underserved communities, disaster prediction systems, multilingual education assistants, government service optimization tools, and emergency response coordinators - all featuring transparent metrics, rights protection, and cultural adaptability.

Footnotes

Based on submissions by: Rural Empowerment and Institutional Development (REPID), Future of Life Institute, SaferAI, Centre for ResponsibleAI, Connected by Data, AI Voices, Data For Good, Digital Scales, GFMD, University of Oxford, and other contributors.

THE FUTURE SOCIETY

Deliverable

3. Launch the AI Commons Initiative to Empower Citizens in AI Design

ℹ️ Context

Current AI development is largely concentrated within the private sector, leading to concerns about transparency, accountability, and the potential for bias and harm. Citizens, especially those from marginalized communities, often lack opportunities to influence the development and use of AI systems that impact their lives

Deliverable 🎯

Launch an "**AI Commons**" initiative to democratize AI by empowering citizens globally to participate in shaping its development and utilization.

- Create **Citizens' Design Councils (CDCs)** to review and provide feedback on AI designs, ensuring that they meet the needs and address the concerns of diverse communities.
- Implement a **Multi-stakeholder Oversight System (MOS)** to monitor AI development and deployment, promoting transparency and accountability.

Timeline 🗓️

Pre-Summit

Develop a detailed framework for the AI Commons, including pilot programs for each pillar.

During the Summit

Announce the AI Commons initiative and invite commitments from countries and organizations.

Future Summits

Begin implementing pilot programs, focusing on convening the first Citizens' Design Council.

Expected benefits ⭐

- ⬆️ Increased **Citizen Participation** in AI Governance
- 🤝 More **Inclusive and Equitable** AI Development
- ✨ Enhanced **Public Trust in AI**

Footnotes

Inspired by submissions from: Missions Publiques, Connected by Data, Forum on Information and Democracy, AI Voices, #Leplusimportant, OpenMined Foundation, Lagori Collective, Data For Good, La Concorde Aula, Renaissance Numérique, Global Coalition for Tech Justice, Open Markets Institute, Mozilla.

THE FUTURE SOCIETY

Insights

1. AI Auditing

🔎 Guiding questions

- What are different approaches to AI auditing and how should these be implemented in order to be taken up at scale?
- What building blocks may be missing to make the market for AI auditing (e.g. certification schemes, training, standardization)?
- What types of other auditing parallels in history might be taken as examples?

🎯 Core Audit Types

Three complementary approaches ensure comprehensive AI oversight:

1. **Internal Audits** - Companies self-assess against their policies
2. **External Audits** - Independent evaluators provide unbiased reviews
3. **Regulatory Audits** - Government-required compliance checks

🏗️ Building Blocks for Success

Four essential elements needed to scale AI auditing globally:

1. **Harmonized Global Standards** - Common international metrics for fairness and transparency
2. **Expert Training Programs** - Specialized education in AI auditing
3. **Technical Tools** - Open-source software and standard documentation methods
4. **Clear Regulations** - Enforceable guidelines for audit practices

📚 Learning From Experience

Key lessons from established fields show us what works:

- **Financial accounting standards** (GAAP) demonstrate how to create transparency
- **Environmental assessments** show effective risk evaluation methods
- **Cybersecurity frameworks** prove the value of continuous monitoring

🎯 Key Takeaway

A robust AI auditing ecosystem requires coordinated global action across technical standards, professional training, and regulatory frameworks, building on proven approaches from other sectors.

Footnotes:

Inspired by the submissions of: Rural Empowerment and Institutional Development (REPID), DTFTP, Safe AI Forum, SaferAI, Centre for Responsible AI (IIT-Madras), Future of Life Institute, Ada Lovelace Institute, AI Voices, The Shift Project, Intermobility.

THE FUTURE SOCIETY

Insights

2. Defining public interest

🔎 Guiding question

- The past decade of work on AI has shown that the term “public interest” encompasses many aspects, including accountability, social justice, human rights, consumer protection, antitrust tools, refusal and curtailment of applications, environmental justice, audits, redressal mechanisms, among others. What existing definitions of ‘public interest AI’ would enable the broader field to come to shared values, goals and outcomes?

🗝️ Key Concepts & Values

Public Interest AI (PIAI) represents artificial intelligence systems designed to serve societal well-being while minimizing potential harms. Core values include:

1. **Transparency and accountability** in development and deployment
2. **Inclusive and equitable** access across communities
3. **Participatory design** involving diverse stakeholders
4. **Alignment** with human rights and environmental sustainability

🏗️ Framework Components

Essential elements for PIAI implementation:

1. **Public Justification**: Clear articulation of societal benefits
2. **Equity Assurance**: Fair distribution of benefits and risks
3. **Citizen Participation**: Active involvement in design and deployment
4. **Technical Robustness**: Security and accuracy guarantees
5. **Validation Access**: Third-party audit capabilities

📋 Governance & Oversight

Key considerations for effective PIAI management:

- Regulatory frameworks balancing innovation and public safety
- Democratic governance processes
- International cooperation and standards harmonization
- Multistakeholder involvement in decision-making

⚠️ Implementation Challenges

Critical obstacles requiring attention:

1. Cultural variations in interpreting AI principles
2. Need for unified ethical standards across borders
3. Political dimensions of AI deployment decisions
4. Balancing private sector innovation with public good

📝 Recommendations

Priority actions for advancing PIAI:

1. Establish clear metrics for public benefit assessment
2. Develop cross-cultural frameworks for ethical AI
3. Create inclusive governance mechanisms
4. Foster international collaboration

Footnotes:

Inspired by the submissions of: Rural Empowerment and Institutional Development, SaferAI, AI Voices, Global Forum for Media Development, DTFTP, Future of Life Institute, Centre for ResponsibleAI (IIT-Madras), Connected by Data, Data For Good, Open Markets Institute, Impact AI, The Shift Project, University of Oxford, Digital Action.

THE FUTURE SOCIETY

Insights

3. Openness

🔎 Guiding question

- What are existing efforts to define what openness in AI means, led by whom, bringing together which stakeholders and across which regions?

🔍 Current Landscape

Several organizations are working to define what openness in AI means. These efforts involve diverse stakeholders from academia, civil society, industry, and government, across various regions, particularly in Europe and North America.

🚀 Key Initiatives

- **UNESCO** and the **OECD** are bringing together stakeholders to develop ethical principles for openness and fairness in AI, focusing on Europe and North America.
- **Mozilla** and the **Columbia Institute of Global Politics** collaborated to create a framework on openness in AI, involving over 40 scholars and practitioners from various regions.
- The **Open Future** initiative is exploring the intersection of openness and AI, aiming to preserve openness to increase transparency and trust. This initiative has a global reach.
- The **Alan Turing Institute** is working to co-create a definition and practice of open AI that meets the needs of diverse communities.

💡 Core Themes

These initiatives highlight a multifaceted understanding of openness in AI:

- **Transparency**: It is crucial to make sure that AI is transparent and based on quality data.
- **Ethical Frameworks**: UNESCO and the OECD are focusing on developing ethical principles for openness and fairness in AI.
- **Open-Source Code and Data**: Mozilla's framework aims to provide a comprehensive understanding of how each component of the AI model stack contributes to openness.
- **Accessible Data Governance Models**: The Open Future initiative examines data governance, transparency, and regulatory issues related to AI.

Footnotes:

Inspired by the submissions of: DTFTP, Rural Empowerment and Institutional Development (REPID), Global Forum for Media Development, Open Data Charter, Concordia AI, University of Oxford, Digital Action, Handicap International.

THE FUTURE SOCIETY

Insights

4. Collective data governance

🔎 Guiding question

- What are examples of approaches to collective data governance (including but not limited to data trusts) that have a proven impact track record of implementation and impact, and may be applicable to the training of AI models (with a specific focus on small models)?

Examples of Successful Approaches:

🤝 Data Trusts:

Data trusts, like the **MIDATA Cooperative** in Switzerland, empower individuals to control and share their health data for research under strict ethical guidelines and governance frameworks. This model ensures data is used responsibly while benefiting both individuals and research initiatives.

🌾 Data Cooperatives:

The **Source Cooperative** from Radiant Earth Foundation is an example of a data cooperative that promotes decentralized data sharing in agriculture. This model allows for data pooling and collective decision-making regarding its use.

🔒 Secure Data Access Initiatives:

The **UK Biobank** offers secure access to a large health database for research while safeguarding privacy. This model demonstrates how sensitive data can be used for AI training without compromising individual privacy.

🌐 European Data Spaces:

Initiatives like **Gaia-X** foster interoperable data spaces in Europe for secure data sharing, promoting ethical use of data for AI training. This model encourages collaboration while maintaining data sovereignty.

Key Considerations:

- 📈 **Transparency and Accountability:** All these approaches emphasize transparent and accountable data governance, ensuring individuals understand how their data is used and who is responsible for its management.
- 🛡️ **Privacy and Security:** These models prioritize robust privacy and security measures to protect sensitive data and maintain public trust.
- 👥 **Community Engagement:** Successful approaches often involve stakeholder participation, ensuring diverse voices contribute to data governance decisions.

Footnotes:

Inspired by the submissions of: Rural Empowerment and Institutional Development (REPID), Open Data Charter, Future of Life Institute, DTFTP, The Shift Project, Open Markets Institute, Saskatchewan Indigenous Cultural Centre, Les e-novateurs.

THE FUTURE SOCIETY

Insights

5. Sustainable energy procurement

🔎 Guiding question

- What are sustainable approaches to procure the energy that power AI training and inference (with a specific focus on small models)? Who is investing in them and how?

🌞 **Renewable Energy Integration:**

1. Some companies, such as Google, have committed to using **100% renewable energy** and are investing in green infrastructure to power their AI operations. ♻️
2. Microsoft is exploring AI solutions to optimize renewable energy source efficiency and data center energy use. 💨

⚡ **Energy-Efficient Infrastructure:**

- Chip manufacturers like ARM and Qualcomm are developing **low-power chips** designed for AI tasks on smaller devices, which can promote local and decentralized energy use.
- **Data center cooling technologies**, such as ambient air cooling (Facebook in Sweden) and experimental underwater data centers (Microsoft), are being explored to reduce energy consumption.
- Intel is working on modular, **energy-efficient AI hardware** that minimizes embodied carbon in server components and can adjust energy consumption based on availability.

🧠 **Optimized Model Training:**

- Organizations like OpenAI are focusing on developing more efficient training methods and energy-efficient models to reduce computational power requirements and energy consumption.

🌍 **Decentralized and Edge Computing:**

- Decentralization through edge computing and TinyML allows small AI models to run on local devices, potentially reducing the need for energy-intensive data transfers to central servers.

While these approaches show promise, Experts highlight that their effectiveness and scalability in significantly reducing the environmental impact of AI remain to be seen. Continued research, development, and collaboration will be necessary to ensure that AI's transformative potential is realized in an environmentally sustainable manner.

Footnotes:

Inspired by submissions from: Rural Empowerment and Institutional Development (REPID), Founder Family, Impact AI, DTFTP, SaferAI, Future of Life Institute, Data for Good, MIT.

04.

Future of Work

60 expert and civil society organizations submissions

AI ACTION SUMMIT

THE FUTURE SOCIETY

Future of Work

EXPERTS' TOP PRIORITIES FOR THE THEME	EXAMPLES OF POTENTIAL KEY DELIVERABLES
1. **Improved Global AI Literacy and Awareness of Risks and Opportunities**	1. **Create a Global AI Education and Critical Thinking Initiative**
2. **Workforce Policies for AI Transition and Job Protection**	2. **Form an International Task Force on AI and Labor Market Disruption**
3. **AI-Specific Educational and Training Programs**	3. **Establish a Global AI Talent Program for the Global South**

THE FUTURE SOCIETY

Deliverable

1. Create a Global AI Education and Critical Thinking Initiative

ℹ️ Context

The increasing ubiquity of AI across industries highlights two interconnected challenges. First, a growing skills gap in AI education and literacy across existing education systems. Second, unequal access to AI resources and expertise, particularly in the Global South, which perpetuates global power imbalances.

Deliverable 🎯

1. **Educational Infrastructure Development**
   - Develop **standardized, multilingual AI curricula for all education levels** (primary through higher education)
   - Create a mobile learning platform available in 6 languages covering AI basics, opportunities, and risks
   - Establish a network of "AI Equity Labs" across the Global South, prioritizing regions with limited AI infrastructure
2. **Implementation Strategy**
   - Deploy comprehensive teacher training and professional development programs
   - Launch regional AI Equity Labs with specialized training programs addressing local contexts
   - Partner with governments to integrate AI skills training into national education systems

Timeline 🗓️

Before the Summit

- Form expert working group
- Develop pilot curriculum and mobile platform (English/French)
- Identify 3 potential AI Equity Lab locations
- Draft funding framework

During the Summit

- Launch platform beta
- Announce first 3 Lab locations
- Secure key partnerships
- Form localization teams

After the Summit

- Open first AI Equity Lab
- Add 2 languages to platform
- Train initial 500 participants
- Launch teacher training program
- Target: Train 10,000 citizens in AI basics and ethics

Expected benefits ⭐

- 🌏 Create a **globally skilled workforce** prepared for AI-driven economies
- 📚 Reduce technological unemployment risk through proactive education
- 🤝 Foster responsible AI development through improved global understanding
- 🏗️ Build local AI educator networks in underserved regions
- 📢 Promote inclusive AI development addressing diverse regional needs

Footnotes

Inspired by submissions from: AI Voices, Global Coalition for Tech Justice, Learning Lab for Resiliency, Centre for ResponsibleAI, Digital Action, Rural Empowerment and Institutional Development (REPID), Concordia AI, AIIC France, AIIC Science Hub, DTFTP.

THE FUTURE SOCIETY

Deliverable

2. Form an International Task Force on AI and Labor Market Disruption

ℹ️ Context

The potential for AI-driven job displacement and labor market disruption requires a proactive and coordinated international response to mitigate negative social and economic consequences.

Deliverable 🎯

Establish a multi-stakeholder task force to develop:

1) **Worker Protection Toolkit**
- Impact assessment template (job quality, autonomy, discrimination risks)
- Human oversight guidelines for AI systems
- Worker wellbeing checklist
- Supply chain labor standards

2) **Collective Bargaining Toolkit**
- Model union agreements (based on Volkswagen/IBM examples)
- Committee formation guidelines
- AI transparency requirements
- Retraining guarantee templates

Timeline 🗓️

Before the Summit
- Form task force with strong union and civil society representation
- Draft initial Worker Protection Toolkit
- Consult with labor organizations

During the Summit
- Launch task force
- Present draft Worker Protection Toolkit
- Demo practical tools
- Gather feedback from stakeholders

After the Summit
- Share first collective bargaining toolkit
- Document early implementation lessons
- Aim to build an International Observatory on AI Impact on the Workforce

Expected benefits ⭐

🌎 Provides a platform for **international cooperation and knowledge sharing**, facilitating a coordinated global response to AI-driven labor market challenges.

🤝 Helps governments and stakeholders **anticipate and adapt to future changes**, minimizing job displacement and fostering a more equitable transition.

✨ Supports the **development of effective policies and programs** to equip workers with the skills needed for the jobs of the future.

Footnotes

Inspired by submissions from: Global Center on AI Governance, DTFTP, Open Community, Impact AI, AI Voices, Rural Empowerment and Institutional Development (REPID), SKEMA Business School, The Center for Information Technology Policy and Workers Algorithm Observatory, The Tony Blair Institute for Global Change.

THE FUTURE SOCIETY

Deliverable

3. Establish a Global AI Talent Program for the Global South

ℹ️ Context

The increasing integration of AI into various industries necessitates a skilled workforce capable of developing, implementing, and managing these technologies responsibly. However, access to AI education and training remains unevenly distributed globally. The Global South faces significant challenges in developing AI talent due to limited resources, infrastructure, and educational opportunities. This gap risks exacerbating existing inequalities and hindering the potential of AI to contribute to sustainable development in these regions.

Deliverable 🎯

- Provide accessible **AI education and training** resources tailored to the specific needs and contexts of the Global South.
- Providing support for existing efforts such as mentorship schemes, fellowships, or other **capacity-building** efforts.
- **Provide broad learning opportunities** for **technical understanding of AI**, as well as **critical thinking** relating to its use, policy implications, risks arising from use-cases of advanced AI, etc.
- **Establish scholarships, funding mechanisms or technology transfer contracts** to support individuals from underrepresented communities in pursuing **upskilling in AI education and careers**.

Timeline 🗓️

Before the Summit

Form a Steering Committee, composed of influential representatives from governments, international organizations, leading AI companies, and prominent academic institutions

During the Summit

Announce the program and leverage the event to secure concrete commitments from stakeholders, encompassing financial support, educational resources, expert mentorship, and future collaborations

After the Summit

Convene the Steering Committee's first official meeting to formalize the program's structure and define immediate next steps.

Expected benefits ⭐

- 🤝 A more **inclusive and diverse global AI ecosystem** that benefits from the contributions of talent from all regions of the world.
- 💡 **Empowered individuals in the Global South** equipped with the skills and knowledge to harness the power of AI for economic and social development.
- 🟰 **Reduced inequalities and increased opportunities for individuals** from underrepresented communities to participate in the AI revolution.
- 🤖 A pipeline of skilled **AI professionals** to meet the growing demand for AI talent in both the Global South and globally.

Footnotes

Inspired by submissions from: Global Center on AI Governance, Australian National University, RadicalxChange Foundation, The Tony Blair Institute for Global Change, Concordia AI, EthicsNet, DTFTP, Future of Life Institute, CIGI, AI Voices

THE FUTURE SOCIETY

Insights

1. Job quality

🔎 Guiding question

- How to leverage AI to increase job quality?

🤖 Automating Repetitive Tasks

A recurring theme is the potential of AI to **automate mundane and repetitive tasks**, thereby freeing workers to engage in more meaningful and fulfilling work. This not only **increases productivity but also enhances job satisfaction**. For example, AI can handle tasks like data entry, scheduling, or routine decision-making.

🛡️ Improving Workplace Safety and Wellbeing

AI can play a crucial role in creating a safer and healthier work environment. AI-powered systems can **monitor workplaces in real-time to detect potential hazards and predict equipment failures**. Additionally, AI can contribute to better work-life balance by enabling **flexible work schedules** and providing tools for **mental health support**.

📚 Upskilling and Reskilling

Respondents emphasize the importance of AI-powered personalized learning platforms to address skill gaps and equip workers for an evolving job market. By analyzing individual performance data, AI can **identify skill gaps and recommend targeted training programs**. This approach ensures workers develop relevant skills, enhancing job satisfaction and productivity.

🌐 Ethical Considerations and International Cooperation

To realize these benefits, ethical standards must be in place to **mitigate risks like bias and inequality**. For instance, algorithms used in **hiring or HR management should be transparent and non-discriminatory**. International cooperation is also crucial for scaling successful efforts and driving global improvements in job quality.

Footnotes:

Inspired by submissions from: Rural Empowerment and Institutional Development (REPID), DTFTP, Open Community, Connected by Data, AI Voices, #Leplusimportant, AI 4 Development Agency (AI4DA), ESSEC Business School & Metalab for Data, Technology & Society, RMIT, Digital Scales, The Tony Blair Institute for Global Change, Impact AI, AIIC France and AIIC Science Hub AI Workstream, University of Geneva, CFE-CGC, Learning Lab for Resiliency®, TRAIL (Trusted AI Labs), Federal University of Rio de Janeiro, Federal University of Amazonas - Institute of Computing (IComp/UFAM).

THE FUTURE SOCIETY

Insights

2. Collective bargaining

🔎 Guiding question

- What are examples of processes to lead collective bargaining when deploying AI in organizations?

A central theme across responses is the need for **early and continuous involvement of workers and their representatives in AI deployment decisions**. Establishing **inclusive committees** comprised of union representatives, employees from diverse departments, and AI experts can ensure comprehensive representation and address concerns proactively. These **committees should be involved in every stage**, from initial consultations and impact assessments to implementation and monitoring.

🛡️ Transparency and AI Impact Assessments

Transparency in AI systems and decision-making processes is crucial for building trust and enabling meaningful collective bargaining. **Organizations should communicate openly about the purpose, functionality, and potential impact** of AI technologies on jobs and work processes. Conducting thorough **AI impact assessments**, which analyze potential job displacement, changes in responsibilities, and new skill requirements, can inform negotiations and mitigate negative consequences.

🌟 Notable Examples and Agreements

- In Germany, **Volkswagen** negotiated an agreement with unions guaranteeing **no layoffs due to AI implementation and committing to retraining programs** for affected employees.
- **IBM** entered an agreement with UNI Global Union ensuring **transparency** in algorithms used in AI systems, **safeguarding workers' rights** to understand how AI impacts their tasks and **limiting AI's role in decisions about layoffs or promotions** without human oversight.

These examples showcase the potential of collective bargaining to ensure a just and equitable transition in the age of AI.

Footnotes:

Inspired by submissions from: Rural Empowerment and Institutional Development (REPID), DTFTP, Open Community, Connected by Data, AI Voices, Digital Scales, Impact AI, CFE-CGC, ESSEC Business School.

THE FUTURE SOCIETY

Insights

3. Working conditions in AI supply chains

🔎 Guiding question

- How to evaluate and improve the working conditions of workers in the AI supply chain (like data labelers), especially with parallels to international efforts in other industries (like textile)?

🕵️ Independent Audits and Transparency

Regular independent audits, akin to those in the textile industry, are vital to assessing and improving data labelers' working conditions. These audits should cover **wages, working hours, safety, and worker rights**. **Transparency** is equally crucial, requiring companies to disclose information about their data labeling partners and supply chain practices.

🛡️ Health and Safety Protections

Ensuring the well-being of data labelers involves **addressing ergonomic risks, mental health challenges, and exposure to harmful or distressing content**. Borrowing practices from **content moderation industries**, regular mental health assessments and clear safety policies should be standard.

💰 Fair Wages and Benefits

Data labelers often endure **low wages and minimal benefits,** despite their pivotal role in AI development. Drawing from initiatives like the **Fairwork project**, respondents advocate for living wages, timely compensation, and essential benefits such as health insurance and paid leave.

🧵 Learning from the Textile Industry

The textile industry offers valuable lessons through programs like the **ILO's Better Work initiative, which improved conditions via social dialogue and fair practices.** Collaborative approaches involving governments, employers, and workers, as seen in the **Bangladesh Accord**, highlight the importance of **transparency and accountability** in achieving ethical labor practices.

Footnotes:

Inspired by submissions from: Rural Empowerment and Institutional Development (REPID), DTFTP, Open Community, Connected by Data, Digital Scales, Impact AI, ESSEC Business School, AI Voices, Acculturia.

05.

Innovation and Culture

58 expert and civil society organizations submissions

THE FUTURE SOCIETY

Innovation and Culture

EXPERTS' TOP PRIORITIES FOR THE THEME	EXAMPLES OF POTENTIAL KEY DELIVERABLES
1. **Enhanced Reporting and Promoting of Green AI Development**	1. **Announce the "GreenAI Leaderboard" to Track and Incentivize Model Sustainability Improvements**
2. **Ethical Governance and Protection of Cultural Data Rights**	2. **Propose Standards for Cultural Data and AI Intellectual Property Rights**
3. **Global Funding for AI Governance and Research**	3. **Coordinate a Fund for International Research Collaborations on AI with a Focus on Ethics, Safety, and Equity**

THE FUTURE SOCIETY

Deliverable

1. Announce the “GreenAI Leaderboard” to Track and Incentivize Model Sustainability Improvements

ℹ️ Context

The development of increasingly larger AI models is causing unsustainable energy consumption, yet there's no standardized way to measure or compare their environmental impact.

Deliverable 🎯

- Create public real-time dashboard tracking carbon footprint of **top 20 AI models** using standardized metrics
- Develop **API** and **submission system** for model developers to report energy consumption data
- Implement monthly recognition program for **most energy-efficient** models and innovations
- Build **automated integration** with major cloud providers for direct energy reporting

Previous Work & Stakeholders ↩️

- **Green IT**'s existing sustainability metrics
- **TRAIL**'s efficiency benchmarking tools
- **The Shift Project**'s energy assessment frameworks
- **UK Advanced Research & Invention Agency**'s computational tracking
- **The Compute Exchange**'s resource monitoring

Timeline 🗓️

Before the Summit

- Finalize metrics and scoring methodology with partners
- Develop platform and API
- Onboard initial 20 AI companies for beta testing

During the Summit

- Official platform launch with 20+ models
- First "Green Innovation" awards ceremony
- Partner showcase and commitment signing

After the Summit

- Monthly leaderboard updates
- Quarterly efficiency awards
- Expansion to include smaller models

Expected benefits ⭐

- 🛡️ Transparent industry-wide standards for AI energy consumption
- 📢 Market incentives for developing energy-efficient models
- ♻️ Reduced environmental impact of AI development

Footnotes

Inspired by submissions from: Policy Network on AI, PauseAI, TRAIL (Trusted AI Labs), The Shift Project, The Compute Exchange, Concordia AI, Rural Empowerment and Institutional Development (REPID), DTFTP, KETI, Alliance Green IT - AGIT, Future4Care, Resilio.

THE FUTURE SOCIETY

Deliverable

2. Propose Standards for Cultural Data and AI Intellectual Property Rights

ℹ️ Context

Inconsistent cultural data standards, inadequate intellectual property (IP) frameworks, and a lack of centralized platforms hinder fair use, effective AI training, and the ability of cultural creators to protect their rights in the AI era. These gaps limit innovation, complicate access to high-quality cultural data, and expose artists and cultural institutions to exploitation.

Deliverable 🎯

This initiative will address key challenges by:

- **Developing international standards** for cultural data formats, metadata, and annotation to ensure interoperability and usability.
- **Launching a pilot platform for cultural data licensing** with clear terms of use, revenue-sharing mechanisms, and support for data exchange between cultural institutions and AI developers.
- **Creating a global IP framework** through an international task force to address data ownership, licensing, and attribution issues, ensuring artists and cultural institutions retain rights over their creations.

Timeline 🗓️

Before the Summit

Convene technical experts, cultural institutions, artists, and legal professionals to draft standards, develop the pilot platform, and form the IP task force.

During the Summit

Present draft standards, pilot platform, and initial IP framework recommendations at the AI Action Summit.

After the Summit

Finalize and publish cultural data standards, pilot results, and draft IP agreements or model legislation.

Expected benefits ⭐

✊ **Empowers Artists and Cultural Institutions**: Protects intellectual property rights and creates new revenue opportunities for creators.

⚙️ **Standardizes Cultural Data Usage**: Improves interoperability and quality, enhancing AI training effectiveness.

⚖️ **Fosters Responsible AI Innovation**: Balances creator rights with innovation needs, establishing a global cultural data market.

Footnotes

Inspired by submissions from: Syndicat français des artistes interprètes, CNRS, International Federation of Actors, RMIT University, culture Solutions.

THE FUTURE SOCIETY

Deliverable

3. Coordinate a Fund for International Research Collaborations on AI with a Focus on Ethics, Alignment, and Equity

ℹ️ Context

Global AI governance lacks coordinated policies for ethical, safety, and cultural challenges, leaving critical issues underfunded and fragmented.
Note: While France's proposed AI Foundation (€485m/year, Nov 2024) addresses public interest in AI, this deliverable could complement it by focusing on ethics, alignment, and cultural innovation.

Deliverable 🎯

🪙 Design multi-track funding to support:

- Ethics and Alignment: Research on AI ethics and societal impact
- Global Equity and Culture Innovation: Projects developing cultural data markets and preservation, expanding compute access in underserved regions
- Prompt large AI corporations to finance this Fund and start a conversation on a potential "Compute Tax"

⚛️ Develop a Framework:

- Create assessment frameworks for project selection across all tracks
- Ensure strong Global South representation and equitable resource distribution

🤝 Foster Collaboration Across Tracks:

- Shared Platforms: Create unified platforms for sharing research, data, and best practices
  - Example: Global Equity and Culture Innovation - Museums can share and monetize their collections for AI training
  - Example: Ethics and Alignment track - Researchers can share AI evaluation methods and results
- Collaborate with existing initiatives like IA2030Mx in Mexico to localize international research and globalize local findings.

Timeline 🗓️

Before the Summit

- Conduct a landscape analysis of existing international AI ethics collaborations.
- Consult with stakeholders, including potential funders, research partners, and international organizations.

During the Summit

- Launch fund with tracks for ethics, cultural innovation, and equity
- Define priorities and secure commitments from partners

After the Summit

- Finalize structure and launch first calls for proposals
- Establish collaborative platforms and exchange programs

Expected benefits ⭐

- 💡 **Advancement of Knowledge:** Address underfunded areas through collaborative technical and cultural research
- 🌏 **Global Collaboration:** Build international partnerships through shared platforms and standards
- 📢 **Policy Impact:** Inform AI governance frameworks considering technical, ethical and cultural aspects

Footnotes

Inspired by submissions from: Concordia AI, Centre for ResponsibleAI (IIT-Madras), International Science Council, RMIT University, Rural Empowerment and Institutional Development (REPID), Stony Brook University, International Federation of Actors (FIA), culture Solutions, Open Future Foundation, AIIC France and AIIC Science Hub AI Workstream, Syndicat français des artistes interprètes, ESSEC, Australian National University School of Cybernetics, Federal University of Rio de Janeiro, Federal University of Amazonas - Institute of Computing (IComp/UFAM), Princeton University Department of Sociology, Acculturia, Minderoo Centre for Technology and Democracy, Open Markets Institute, Data & Society, Pour Demain.

THE FUTURE SOCIETY

Insights

1. Data on environmental impact of AI

🔎 Guiding question

- What are your views on new data on the environmental impact of AI (trends and scale of the impact compared to the use)?

⚡ Energy-Efficient Algorithms and Hardware

New data highlights the growing **environmental impact of AI,** particularly the **energy-intensive** process of training large models. To combat this, there is a call for **investment in energy-efficient algorithms and hardware**. This includes developing **optimized AI chips and algorithms** that minimize energy consumption while maximizing performance.

♻️ Lifecycle Assessments and Sustainable Infrastructure

Lifecycle assessments are vital to evaluate AI's environmental impact from production to disposal. Drawing on practices from other industries, respondents advocate for sustainable infrastructure and **stronger collaboration** among AI developers, environmental scientists, and policymakers to address this challenge comprehensively.

🌞 Transitioning Data Centers to Renewable Energy

Respondents emphasize the critical need for **transitioning data centers to renewable energy** sources. Solar, wind, and hydroelectric power adoption can be incentivized to reduce AI's reliance on fossil fuels and promote greener operations.

🌍 International Cooperation and Clear Targets

The AI Action Summit has a unique opportunity to lead by setting measurable targets for reducing AI's environmental footprint. **Collaborative frameworks between governments, industry leaders, and researchers are essential** for ensuring that Green AI initiatives are effective and scalable.

Footnotes:

Inspired by submissions from: Rural Empowerment and Institutional Development (REPID), DTFTP, Mohamed Bin Zayed University of Artificial Intelligence, I L Expansions, RMIT University, Open Future Foundation, Lagori Collective, The University of Cambridge Leverhulme Centre for the Future of Intelligence, International Science Council, Acculturia, Minderoo Centre for Technology and Democracy, Open Markets Institute, Fundação Itaú, bny, Princeton University Department of Sociology.

THE FUTURE SOCIETY

Insights

2. Market of cultural data

🔎 Guiding question

- How to organize a market of cultural data to train AIs (technical protocols, economic dynamics, parallels to other markets, nature of the market, ways to attract supply and demand)?

⚙️ Technical Protocols

Standardized data formats and protocols are paramount to ensuring data interoperability and quality. Key components include:

- **Data Standardization:** Establishing common formats and standards for data collection, storage, and sharing.
- **Metadata and Annotation:** Using detailed metadata and annotations to enhance the usability and relevance of cultural data for AI training.

💰 Economic Dynamics

A thriving cultural data market requires an economic model that incentivizes both data providers and consumers:

- **Transparent Platform:** Proposed is a transparent platform where cultural data can be licensed or sold, featuring revenue-sharing models to incentivize data contributors. Dynamic pricing based on data quality and demand could emulate platforms like stock photography or music services.
- **Government Support:** Governments can attract supply by providing grants or subsidies to smaller cultural institutions and creators.

📊 Parallels to Other Markets

Insights from existing markets can inform the structure and operation of cultural data markets:

- **Intellectual Property (IP) Markets:** Lessons from IP licensing systems where creators license their works for use.
- **Digital Content Markets:** Models from digital content markets (e.g., music, video) that monetize content through subscriptions, pay-per-use, or ad-supported systems.

Footnotes:
Inspired by submissions from: Rural Empowerment and Institutional Development (REPID), DTFTP, Mohamed Bin Zayed University of Artificial Intelligence, RMIT University, ESSEC.

THE FUTURE SOCIETY

Insights

3. Cross-border innovation

🔎 Guiding question

- How to generate innovation in AI that is not only focused on national silos?

🤝 Global Partnerships

Forming global partnerships is crucial to dismantling silos. Key strategies include:

- **Joint Research Projects:** Encouraging collaborative research projects that bring together experts from different countries.
- **Shared Funding:** Establishing joint funding mechanisms to support international AI research and development.
- **Collaborative AI Labs:** Creating AI labs across diverse regions to foster knowledge sharing and cross-cultural collaboration.

📂 Open-Source Platforms and Datasets

Promoting **open-source platforms and shared datasets** is essential for democratizing access to AI tools. These initiatives enable contributions from a broader range of countries, accelerating innovation and ensuring that AI benefits a more inclusive global community.

📜 Standardized Regulations

Standardized regulations, particularly around data governance and AI ethics and safety, are critical for cross-border interoperability. Developing a global framework addressing ethical considerations and promoting responsible AI development can enhance international collaboration.

Footnotes:
Inspired by submissions from: Stony Brook University, Open Future Foundation, Mohamed Bin Zayed University of Artificial Intelligence, DTFTP.

Acknowledgments: Contributing Organizations

While submissions from governments and corporations were not accepted, the expert consultation included submissions from leading academic institutions, research institutes, civil society organizations, think tanks, professional associations, foundations, and independent experts. Organizations included: 5Rights Foundation, 7amleh, Acculturia, Active Service for the Benefit of Education and its Reform, Ada Lovelace Institute, Aivancity, AI 4 Development Agency (AI4DA), AI Forensics, AI Safety Asia, AI Voices, AiXist, Alexander von Humboldt Institut für Internet und Gesellschaft, Alliance Green IT - AGIT, Alter.org.il, ANU School of Cybernetics, ARIA, Artificial Impact, Association Green IT, Australian National University, School of Cybernetics, Berggruen Institute, Carnegie Endowment for International Peace, Carnegie Mellon University, CFE-CGC, Centre for International Governance Innovation, Centre for ResponsibleAI (IIT-Madras), Centre pour la Sécurité de l'IA (French Center for AI Safety), Center for a New American Security, Center for Long-Term Cybersecurity (CLTC) at the University of California, Berkeley, CIFAR, CNRS, Confiance.ai, Connected by Data, Concordia AI, Convergence Analysis, Cooperative AI Foundation, Culture Solutions, Data & Society, Data For Good, Data Privacy Brasil, Datasphere Initiative, Digital Action, Digital Scales, DFKI, DTFTP, Effective Institutions Project, EthicsNet, ESSEC Business School & Metalab for Data, Technology & Society, Existential Risk Observatory, FASAP:FO, Federal University of Amazonas - Institute of Computing (IComp/UFAM), Federal University of Rio de Janeiro, Fondation IA pour l'École, Fondation Pierre Fabre, Forum on Information and Democracy, Forseti (Doctrine), Future of Life Institute, George Washington University, Global Center on AI Governance, Global Forum for Media Development (GFMD), Global Partners Digital, Handicap International - Humanity Inclusion, Imec, Impact AI, Institut de Recherche en Éthique du Sujet Numérique (IRESN), Institute for AI Policy and Strategy, Institute for Ethics in Artificial Intelligence, Technical University of Munich, InterAgency Institute and Lusófona University, Intermobility, International Center for Future Generations, International Federation of Actors (FIA), International Science Council, InternetLab, Laboratoire de l'Égalité, Lagori Collective, La Concorde Aula, La Villa Numeris, Learning Lab for Resiliency®, Les E-novateurs, Lincoln Alexander School of Law, METR, Minderoo Centre for Technology and Democracy, Missions Publiques, Mohamed Bin Zayed University of Artificial Intelligence, Mykolas Romeris University, National Society for the Prevention of Cruelty to Children (NSPCC) UK, New York University, Open Community, Open Data Charter, Open Future Foundation, Open Markets Institute, OpenMined Foundation, Oxford Martin AI Governance Initiative, PauseIA, Pour Demain, Princeton University, RadicalxChange Foundation, Renaissance Numérique, Reporters Without Borders (RSF), Resilio, Rick Gillespie, RMIT University, RSA Conference, SaferAI, Santé Publique France, Siasa Place, SKEMA Business School, Société Française des Traducteurs, Stanford, Stony Brook University, Swedish Defence University, Syndicat Français des Artistes Interprètes, The Centre for Long Term Resilience (CLTR), The Compute Exchange, The Safe AI For Children Alliance, The Shift Project, Tony Blair Institute for Global Change, TRAIL (Trusted AI Labs), University of Buenos Aires / School of Law, University of Burgundy, Observatoire de l'éthique publique, Institut Universitaire de France, University of Cambridge Leverhulme Centre for the Future of Intelligence, University of Montreal/Mila, University of Notre Dame Law School, University of Oxford, University of Pennsylvania, University of São Paulo, University of Toronto and Vector Institute, Youth for Privacy, Z Inspection Initiative.

THE FUTURE SOCIETY

CONCLUSION

AI ACTION SUMMIT

Top priorities and recommendations

Lay the foundation for strong, global, multistakeholder AI governance

Global AI Governance

- International Scientific Panel on AI Risks
- Mechanisms for CSOs and Global South Inclusion

Enable access to qualitative AI education and training for everyone

Future of Work

- Global AI Education and Critical Thinking Initiative
- International Task Force on AI and Labor Market Disruption

Establish shared standards for safe, responsible AI

Trust in AI

- Common Thresholds for AI Oversight
- AI Corporation Commitments Report Card

Ensure AI's environmental benefits outweigh its costs

Innovation & Culture

- Mandatory, Auditable Model Efficiency Standards
- Global GreenAI Leaderboard

Prioritize AI solutions addressing existing issues, protecting our rights

Public Interest AI

- Global Charter for Public Interest AI
- AI Commons Initiative to Empower Citizens in AI Design

Key takeaways from the open consultation

A Defining Moment for AI Governance

The 2025 AI Action Summit represents a pivotal opportunity to redefine how the world governs artificial intelligence. Bold commitments from global leaders are essential to address the profound societal challenges posed by AI. This effort must be grounded in collaboration with citizens and experts to design solutions that are inclusive and impactful.

Voices Driving the Vision

Citizens participating in consultations have demanded robust policies to protect their rights and agency from the unchecked deployment of AI systems driven by private interests. They support the use of accountable AI technologies to address pressing global challenges, including medical research, environmental disasters, and public innovation. Additionally, citizens call for universal access to educational resources to ensure that AI's benefits are distributed fairly and society is prepared for the future.

Experts from civil society and academia echo these priorities, emphasizing the need for global standards that ensure fairness, transparency, privacy, and sustainability in AI governance. They advocate for significant investments in research and international cooperation to mitigate long-term risks, including systemic and existential threats. Ensuring equitable access to the benefits of AI is seen as crucial for reducing systemic inequalities and enabling inclusive progress, regardless of origin, location, gender, or income.

A United Call for Bold Leadership

This vision is supported by the insights of over 10,000 citizens and 200 experts gathered through consultations led by ENS-PSL, CNNum, The Future Society, Make.org, and Sciences Po's Tech & Global Affairs Innovation Hub. Together, these voices highlight the need for inclusivity, legitimacy, and measurable impact in shaping the Summit's outcomes.

Seizing the Historic Opportunity

The Summit must rise to the occasion and deliver a comprehensive roadmap for a future where AI serves humanity's interests rather than dictating them. This is a historic moment for bold leadership and meaningful action, with the world watching closely to see what unfolds.

Acknowledgments: Authors

Constance de Leusse, AI & Sociéty Institute (ENS-PSL) and SciencesPo Tech & Global Affairs Innovation Hub
Nicolas Moës, The Future Society
Axel Dauchez, Make.org
Jean Cattan, Conseil National du Numérique
Caroline Jeanmaire, The Future Society
Tereza Zoumpalova, The Future Society
Alexis Prokopiev, Make.org
Marthe Nagels, Make.org
Victor Laymand, Make.org
Pierre Noro, SciencesPo Tech & Global Affairs Innovation Hub
Mai Lynn Miller Nguyen, The Future Society
Niki Iliadis, The Future Society
Jules Kuhn, Make.org

AI ACTION SUMMIT

SciencesPo
Paris School of International Affairs

AI & Society Institute

Fondation Abeona
Championing Responsible AI

Dauphine | PSL

ENS | PSL

THE FUTURE SOCIETY

conseil national du numérique

MAKE.ORG