

# A Global Civic Debate on Governing the Rise of Artificial Intelligence

Report 2018



## Contents:

|                                |       |
|--------------------------------|-------|
| Executive Summary              | p. 4  |
| Our Promise                    | p. 15 |
| Key insights                   | p. 17 |
| Top recommendations            | p. 51 |
| Tales from Imaginaries         | p. 65 |
| Our Methodology                | p. 76 |
| Debate Health                  | p. 78 |
| Beyond the Global Civic Debate | p. 80 |

### Authors:

Nicolas Mialhe  
Arohi Jain  
Caroline Jeanmaire  
Yolanda Lannquist  
Josephine Png

The Future Society © 2018  
[www.thefuturesociety.org](http://www.thefuturesociety.org)

## 5 languages

English, French, Chinese,  
Japanese, Russian

## 17 Global partners:

H5, Bluenove, IEEE, CNIL,  
Les Echos, JSAI Ethics Committee,  
Usbek & Rica, Cap Digital,  
CNNum, World Future Society,  
Tencent Research Institute,  
The Millennium Project  
Future of Life Institute, Centre  
for the Study of Existential Risk,  
CFI, Nesta, GCSP

## 21 Online & Offline Events

## Global reach:

From China, US, Europe and Russia.

Engagement & Participation:

**70 390** Page views  
**2 525** Learning hours  
**2 200** Participants  
**2 074** Votes  
**1 291** Written proposals  
**702** Writers

## 21 Newsletters

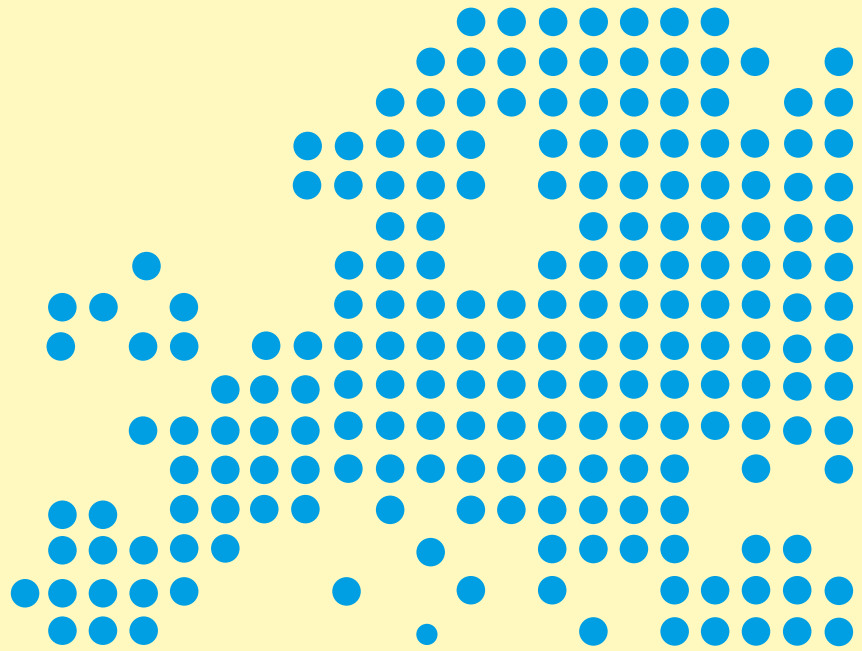
circulated to keep the community informed  
on the latest developments in the discussion

Social Media:

**#AlforGood**  
and **#ImaginariesofAI**  
reached over **5,5** Million



# Executive Summary



Artificial Intelligence and machine learning are becoming irrevocably embedded in our lives. From medical and legal diagnostics to financial loan approvals and virtual personal assistants, AI has many uses that benefit our society through greater inclusion and access to goods and services. It beckons a step-change in humanity's progress, and with continuous innovation, the positive impact of AI could amplify over time. Nonetheless, as with previous technological revolutions, AI also poses significant risks and threats to humanity, such as disintegration of individual privacy, lack of agency and control, scaled threats to cybersecurity, prolonged social disorder, and many others.

The opportunities and challenges of the AI revolution are inextricably imbricated. Designing and deploying smart governance for AI implies striking the right balance between pursuing the upside and minimizing the downside of this technological revolution. Smart governance also must differentiate between that which can be kept culturally specific and that which needs to be universal. This is particularly challenging in a global context that is changing fast, where nations, people, and economies are becoming more interdependent and connected, yet still very distant in terms of culture and values. Emerging economies for instance are driven by the urgency of development, modernization, demography, and disparity, while more advanced countries strive to maintain their position among world powers by reforming and renewing their socio-economic models. Harmonizing value systems in such context is difficult.

How the AI revolution plays out is not destiny or chance. Citizens, governments, business leaders, and academics are empowered to decide the course. This is why we convened the Global Civic Debate, a cross-disciplinary community assembled to discuss the multi-layered aspects of AI and its governance over a period of seven months. Through a diverse collection of more than 2,000 participants, over 700 writers, more than 3,300 contributions in five languages, and 21 events, we garnered insights into the impact of AI and discovered how it reverberates around the world. We synthesized these insights and developed recommendations for policymakers to shape a positive future for the role of AI.



# Key Insights

Based on this collective intelligence exercise, we curated the following key insights.

## The prevailing notion of Artificial Intelligence is obsolete.

1

It does not do justice to the reality (or the dynamics) of the concept it attempts to describe. It is a legacy of a history where there was never a universally accepted definition, but always ambiguity. The notion remains broad, fluid, and contested. AI refers to a vast array of techno-scientific disciplines, technologies, and methods, a loose definition that has accompanied its deployment across numerous sectors, which furthered the ambiguity. While AI technology has evolved, the public remembers film-depicted fantasies of machine intelligence replicating common sense, consciousness, or sentience. For effective governance, the notion must be re-invented, and inconsistencies and inaccuracies must be overcome by understanding the components and drivers of AI. Our functional definition – *big data driven, machine learning algorithm-centric, complex socio-technical systems powered by supercomputing* – serves as a lens and an anchor to assess the current dynamics of AI and how it should be governed.

## Expectations of AI will diverge within and among societies, but building trust in AI is a universal desire.

2

There are no foundational moral, ethical, or societal norms that guide a universal view on an AI-driven future. However, it is evident that creating trust in AI technologies is valuable to the debate community. With continued engagement through public forums and awareness campaigns, technical standards and certification, and greater resources dedicated to explainability, citizens are more likely to gain trust in AI. This is an important hurdle, because when AI is deployed at scale, its full benefits will only be unlocked if citizens trust the technology enough to consistently engage with it in their everyday lives.

# There is potential for blockchain and other self-governing protocols to address challenges in AI.

3

Blockchain and distributed ledgers could be used to address key challenges in AI, such as securing personal data, access to data for more developers, and governance of ethical systems. Blockchains and other self-governing protocols use distributed and decentralized digital ledgers to record, store, and cryptographically secure transaction data between parties. This infrastructure raises potential applications for AI such as securing data and the potential to support a *data exchange*, which may encourage more citizens to offer data, garner greater access to data for more parties, and increase diversity of sources of data reducing data bias. Blockchain-based incentivization mechanisms and smart contracts may also offer complement and/or alternative models for governance of AI development. These potential applications are largely early-stage, unexplored and may face technical challenges, but they offer interesting theoretical complements/ alternatives to addressing the major challenges in AI governance.

# AI is envisioned to unlock a step-change in societal progress.

4

With its vast potential to fuel and support new scientific discoveries, and ability to create impact at scale, such as monitoring, progressing, and achieving the UN Sustainable Development Goals, AI could beckon a clear shift from our current societal structures, which are arguably leaning toward increased inequality and exclusion. With robust global cooperation and deployment of AI, this technology could help address major challenges such as climate change and poverty, and find new or more efficient allocations of resources that drive economic development. Humanity should thrive if the appropriate checks and balances are in place for governing the rise of AI.

Smart governance is key to navigating AI risks, achieved by being agile and evolving.

5

From risks of data privacy to large scale workforce displacement, our debate surveyed a range of potential threats and risks as AI becomes increasingly prevalent. The debate took a clear stance that governance plays an important role in managing risks and ensuring that values such as inclusion, equality, representation, and transparency are preserved. Governance is also key to avoid an industry or national *race to the ethical bottom* to gain market share through first-mover advantages. Given the unprecedented pace of technological change, governance must evolve on several levels, including integration of the local and the global, and alignment of the political and the technical. Smart governance would enable governments to develop better knowledge on these deeply complex topics.

Anxieties over loss of human safety, control, or major failures in AI-integrated systems merit careful consideration by the global AI community.

6

With increased dependence on AI to conduct critical daily decisions – from individual choices such as scheduling time to supranational choices of deploying autonomous weapons – there is a significant fear of the consequences when humans are removed from the command chain. This, coupled with the lack of trust in the AI decision-making processes, creates significant roadblocks to public acceptance and adoption of AI.

AI makes humans re-evaluate their identity and place in society.

7

As the nature of machine intelligence evolves toward general or super intelligence, the shifting paradigm of human beings, the most cognitively developed entities so far known in the Universe, necessitates a discussion on the relevance, role, and future development of humanity in an AI centric world.

.0. 00000 88 00000 000 080  
888. `888' .8' `888' `8' `"  
"888. 888 .8' 888 8 000. .00. 0000 0000 000 .00000. 0000 d8b .0000.0 .000  
`888. 888 .8' 888 8 `888P"Y88b `888 `88. .8' d88' `88b `888""8P d88( "8 `P  
008888. 888 .8' 888 8 888 888 888 `88..8' 888000888 888 `Y88b. .oP  
`888. 888 .8' `88. .8' 888 888 888 888 `888' 888 .o 888 o. )88b d8(  
o8888o o888o 88 `YbodP' o888o o888o o888o `8' `Y8bod8P' d888b 8""888P' `Y88

000. .o8 0000 . 88 0.000000..0  
`Y8b "888 `888 .o8 d8P' `Y8  
888 8880000. .0000.0 .00000. 888 .00000. .o88800 .00000. Y88bo. .00000  
888 d88' `88b d88( "8 d88' `88b 888 d88' `88b 888 d88' `88b `Y8888o. d88' `88b  
888 888 888 `Y88b. 888 888 888 888000888 888 888000888 `Y88b 888000888  
d88' 888 888 o. )88b 888 888 888 888 .o 888 . 888 .o oo .d8P 888 .o  
od8P' `Y8bod8P' 8""888P' `Y8bod8P' o888o `Y8bod8P' "888" `Y8bod8P' 88 8""88888P' `Y8bod8P'

000. 0000 .o8 .  
`Y88. `888 "888 .o8  
.d88' .00000. 0000 d8b .0000.0 .00000. 000. .00. .0000. 888 .0000888 .0000. .o88800 .0  
o88P' d88' `88b `888""8P d88( "8 d88' `88b `888P"Y88b `P )88b 888 d88' `888 `P )88b 888  
888000888 888 `Y88b. 888 888 888 888 .oP"888 888 888 888 .oP"888 888  
888 .o 888 o. )88b 888 888 888 d8( 888 888 888 888 888 d8( 888 888 . d8(  
`Y8bod8P' d888b 8""888P' `Y8bod8P' o888o o888o `Y888""8o o888o `Y8bod88P" `Y888""8o "888" `Y8

0000. 0000 0000 0000 080 88  
`Y8b `888 `888 .00000. .00000. 888 0000 .00000. 888 .00. .0000. 0000 000. .00. 88  
00888' 888 d88' `88b d88' `Y8 888 .8P' d88' `Y8 888P"Y88b `P )88b `888 `888P"Y88b .8'  
`88b 888 888 888 888 888888. 888 888 888 .oP"888 888 888 888 .8'  
.88P 888 888 888 888 .o8 888 `88b. 888 .o8 888 888 d8( 888 888 888 888 .8'  
od8P' o888o `Y8bod8P' `Y8bod8P' o888o o888o `Y8bod8P' o888o o888o `Y888""8o o888o o888o o888o 88

000. 88 0.000000..0  
`Y88. .8' d8P' `Y8  
.d88' 0000 d8b .00000. .00000000 0000 d8b .00000. .0000.0 .0000.0 Y88bo. 000. .  
o88P' `888""8P d88' `88b 888' `88b `888""8P d88' `88b d88( "8 d88( "8 `Y8888o. `888P"  
888 888 888 888 888 888 888000888 `Y88b. `Y88b. `Y88b 888  
888 888 888 `88bod8P' 888 888 .o o. )88b o. )88b oo .d8P 888  
d888b `Y8bod8P' `8000000. d888b `Y8bod8P' 8""888P' 8""888P' 88 8""88888P' o888o

000. 88  
`Y8b .8' d8P' `Y8  
.00000. 0000 000 .00000. 0000 d8b 000. .00. .0000. 000. .00. .00000. .00000.  
d88' `88b `88. .8' d88' `88b `888""8P `888P"Y88b `P )88b `888P"Y88b d88' `Y8 d88' `88b  
00000 888 888 `88..8' 888000888 888 888 888 .oP"888 888 888 888 888 888000888 .  
.88' 888 888 `888' 888 .o 888 888 888 d8( 888 888 888 888 .o8 888 .o .8  
od8P' `Y8bod8P' `8' `Y8bod8P' d888b o888o o888o `Y888""8o o888o o888o `Y8bod8P' `Y8bod8P' 88

.0. 080 0000 .o8 000000000000  
888. `"  
"888. .00000000 0000 888 .00000. "888  
`888. 888' `88b `888 888 d88' `88b `P )88b `888P"Y88b d88' `888 888 0000 000  
008888. 888 888 888 888 888000888 .oP"888 888 888 888 888 888 8880008 88. .8'  
`888. `88bod8P' 888 888 888 d8( 888 888 888 888 888 888 888 " `88..8'  
o8888o `8000000. o888o o888o `Y8bod8P' `Y888""8o o888o o888o `Y8bod88P" o88800000d8 `88'  
d" YD  
"Y88888P'

.o8 .o80 . 88 00000 00000  
"888 .o8 `"  
.0000888 .00000. 000. .00. .o88800 0000 .o88800 0000 000  
d88' `888 d88' `88b `888P"Y88b 888 `888 888 `88. .8'  
888 888 888000888 888 888 888 888 `88..8'  
888 888 888 .o 888 888 888 . 888 888 . `888'  
`Y8bod88P" `Y8bod8P' o888o o888o "888" o888o "888" .8' 88  
.o..P'  
`Y8P'

# Top recommendations

To address the insights outlined above, the debate community put forward the following high-level recommendations.

## Build *decision intelligence engineering* capability into governance processes to achieve balance in policy making to meet societal goals.

1

Decision intelligence builds an iterative understanding of the way certain actions lead to an outcome. The engineering concept relates to understanding who has the skill set to handle a particular challenge. In AI, decision intelligence engineering enables the right mix of talent to solve a problem and ensure the results are as intended. To disentangle the complexity in AI technologies and adequately address their impact, such an iterative and practical mechanism is essential for effective governance. Our decision intelligence capability could be advanced through these concepts:

- An Intergovernmental Panel on AI (IPAI) to build a solid base of facts to enable policy making. A large and interdisciplinary group of AI scientists and experts who perform regular assessments nested in a solid scientific process could play a key role in forging global consensus on the challenges to be addressed in AI development.
- An *AI4SDGs Center*, which seeks to harness the power of AI to monitor, simulate, predict, and make progress toward the UN's Sustainable Development Goals. The center – a *do tank* that brings together business, governments, academia, and civic society – would act as an engine to apply IPAI's policy recommendations in real-world projects, embed governance and ethics in AI problem solving, and pursue inclusive and diverse AI development.
- A move from *United States versus China* toward *United States with China*, to close the gap in their AI business models, technical standards, and underlying framings and outlooks. The U.S. and China increasingly appear as *AI superpowers* operating within a global duopoly. Exploring the appropriate articulation between competitive and collaborative dynamics seems an important endeavor.

# Propagate effective global governance for AI.

## 2

- Embed a multi-stakeholder, inter-disciplinary and collective intelligence process for policy making and oversight. This could include bringing artists from the film industry into conversations around governance, as they offer valuable insights and influence over citizens.
- Favor the rise of international technical standards-setting processes and organizations for AI to harness the power of *soft governance*.
- Work with established organizations and communities of practice to build and operationalize universal but culturally adaptive ethical guidelines, codes of conduct, and codes of practice.
- Explore the need for international treaties and implementation and oversight mechanisms (*hard governance*) on several topics. This could include governing the collection, storage, processing, and flow of data as well as the testing, liability, and use of autonomous drones, and the deployment of lethal autonomous weapons.

# Ensure representative and secure data for AI.

## 3

- Establish a global data exchange that serves as an independent, market-driven, regulated platform for data exchange.
- Leverage development in blockchain technologies to secure data from its source.
- Educate citizens about AI to boost data representation, inclusion, and civic engagement.
- Build mechanisms whereby AI is able to detect and correct its own level of bias or lack of representation.

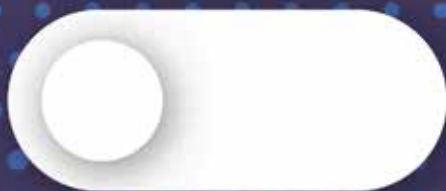
4

- [illegible]

# Beyond the Global Civic Debate

The Global Civic Debate began a conversation on the rise, dynamics, and consequences of AI. While this report provides an overview of the main insights, tensions, and policy recommendations put forward by our community, we recognize this is just the beginning. The Future Society commits to continuing this endeavor in the following ways:

- Work closely with governments and policymakers through research, advisory, convenings, educational programs, and *special projects* to ensure our community's views are translated into actionable policy and impact.
- Continue to nurture the debate and engage a diverse and inclusive community by building grassroots movements. We plan to disseminate many of the insights from the debate to global citizens through engaging video and visual content that spreads awareness and makes the debate on AI increasingly inclusive.



- Harness the creativity in the imaginaries of AI as a tool for communicating and working with governments, business leaders, and citizens. Among others, we aim to work with script writers, producers and others in the entertainment industry to understand how imaginary dynamics impact public perception and thus governance processes. We aim to create sharper and more balanced depictions of AI for mass scale audiences.
- Continue to leverage the collective intelligence platform to bridge conversations on AI among government, academia, private sector, and civil society. Inspired by the Imaginaries of AI theme of the Global Civic Debate, we are developing methodologies and strategies for large organizations (public and private) to better understand and harness the impact of AI adoption, implement effective AI solutions within their operations, and enable transformation.
- Work to develop pathways towards Smart Global AI Governance through a virtuous cycle of research, convenings, and education programs. We are committed to applying a dynamic and iterative approach informed by the latest collective intelligence tools and methodologies to better understand the dynamics, impact and governance of AI. One example is the *“Dubai Global Governance of AI Roundtable”* (GGAR), which we are developing in partnership with the United Arab Emirates government and which is envisioned as a yearly event fueled by a continuous research and community building effort. The GGAR first edition took place in Dubai last February during the 2018 World Government Summit. Another example is our support for the creation of an *“Intergovernmental Panel on AI”* (IPAI; modelled after the Intergovernmental Panel on Climate Change) envisioned as a vessel to mature a solid base of matter of facts capable of informing policy-making in the long run. A final example is our proposal to develop an *“AI4SDGs Center”*, envisioned as an open platform to innovate, test and help scale novel AI solutions and associated models of Public-Private-People Partnerships and ethical governance to accelerate the delivery of the 17 *Sustainable Development Goals* (SDGs).

Consistent with the nature of AI, we believe our thinking on the impact and consequences of the AI revolution will continually evolve. We view this first Global Civic Debate as a robust foundation to continue the work and help build a future where we can capture the upsides of AI while minimizing the downsides risks and effects.

# Our Promise

The advent of Artificial Intelligence and its widespread application will create profound and irrevocable changes to humanity. This technological surge is hailed as the most revolutionary in history because it is expected to transform the fabric of our society, cutting across civilizations, industry domains, and national boundaries. Understanding the depth of this impact remains nascent, with no one country, industry, discipline, or group of experts having the capability to holistically manage the revolution alone.

To address this challenge, in September 2017 we launched a Global Civic Debate to understand the impact and governance of the rise of AI. We led this endeavor through a curated collective intelligence platform and methodology, as well as online and offline events, which were inclusive and open to everyone.

Many people, from media to academics and experts to business leaders, were calling for a public debate on how the AI revolution should be governed. Through grassroot efforts and an array of global partnerships, we convened a diverse community to better understand how the revolution is playing out, its consequences, and how they must be governed to maximize benefits, minimize risks, and ensure the benefits reach everyone.

Our civic debate confronted challenging questions such as: How can we use AI to increase access to credit, insurance, or personalized medicine, without sacrificing our privacy and security? How do we revamp our labor markets and education systems to deal with the massive automation wave? How do we prevent an arms race or ensure these systems are safe and benefit humanity? These are big questions that will impact us all. Since no single entity can approach such questions alone, we deployed the collective intelligence of a diverse community to search for solutions. It is our conviction that the AI revolution is a global phenomenon which calls for global coordination. We therefore engineered a global conversation by making the platform available in five languages (English, French, Chinese, Japanese, and Russian), which transcended civilizational boundaries. We sought to design a discussion that had the courage to confront and align value systems. We knew this would be a challenge but pursued this ambition with citizens across the world.

In closing the civic debate in March 2018, we were deeply encouraged by the progress made, not only in answering the questions but also in meaningfully engaging with the community to unearth salient insights regarding our hopes, fears, and expectations. This report is a first effort to harness and synthesize the deluge of data gathered from our multilingual community. In the following sections, we present the key insights that have resulted from this cross-country collective intelligence exercise, as well as practical policy recommendations.

We recognize there is much more to be done to harvest the breadth of contributions from our community, going deeper into each discussion theme and discovering new policy solutions. This collective intelligence exercise was a first step to nourish global thinking on AI and foster inclusivity. We will continue to rely upon our community's insights to shape discussion and public policy on AI that is representative, inclusive, and protects our collective values.

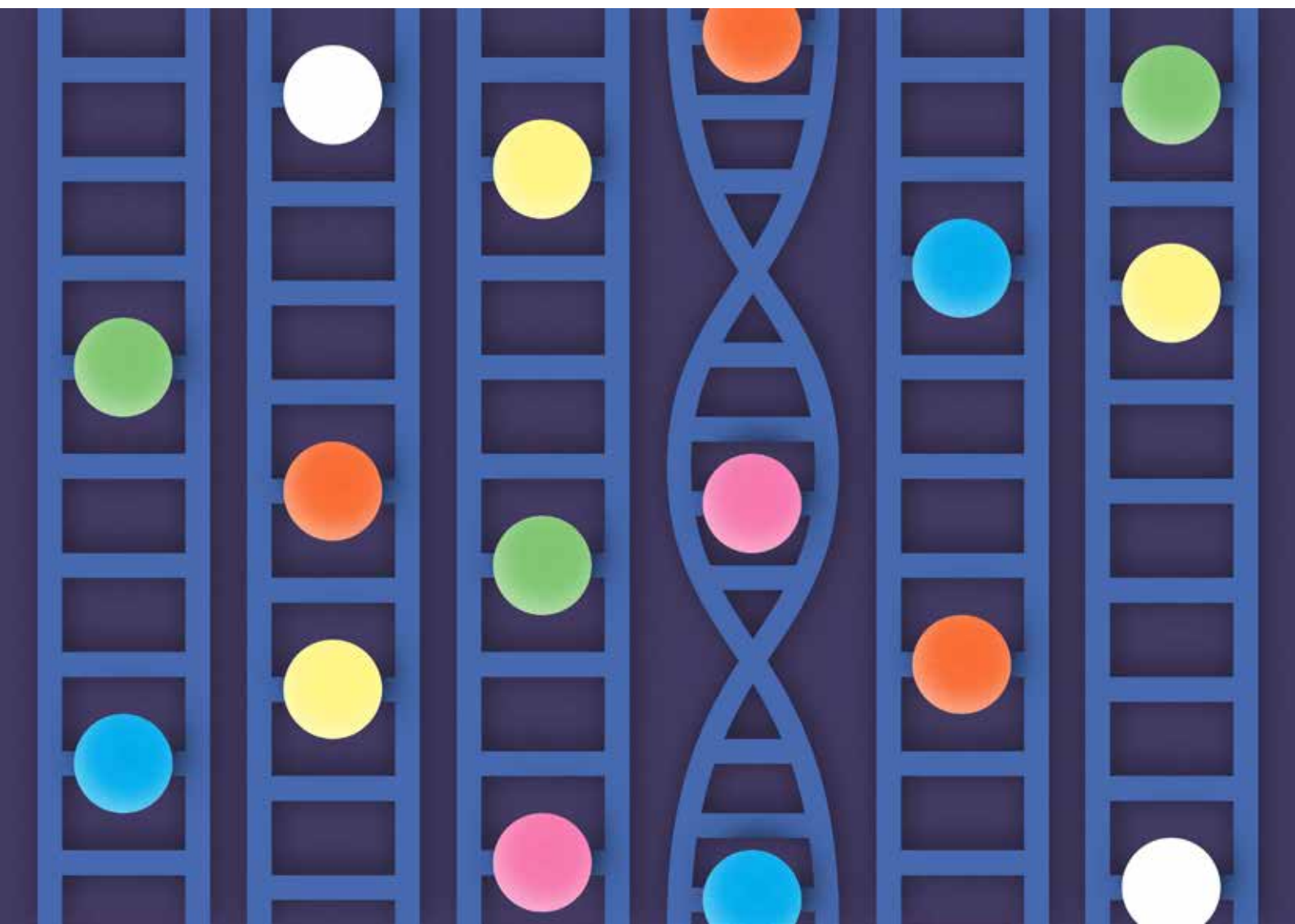
# Key Insights



# 1 The evolving notion of AI

As technological development and deployment progress, both humans and AI will experience shifting identities and roles. The notion of AI will continue to transform alongside changing societal, ethical, and legal norms. Moving beyond human-imitative and toward intelligent infrastructure, AI is likely to take on roles in managing complex systems, such as transportation, energy, supply chains, and potentially political or economic systems.

As the impact of AI becomes increasingly pervasive across greater domains, our society will experience significant transformation.



## The prevailing notion of AI is obsolete

There is no universal definition of AI; the boundary to define it is highly contested, evolving, and varied. Despite its status as the most profound techno-scientific revolution of our time, AI maintains significant ambiguity in its connotation, which is broad and fluid. Today, AI refers to a vast array of technologies and methods, including robotics, biometrics, swarm intelligence, virtual agents, natural language generation, semantic technology, computer vision, machine learning, and more. A loose definition of AI has accompanied its deployment across numerous domains and sectors, which in turn furthered the ambiguity.



“Regulation for AI is misleading.  
The technologies are so different –  
we have to stop calling everything AI.”

As the concept of AI evolves, some note John McCarthy’s dictum :  
*“As soon as it works, no one calls it AI anymore.”* An odd paradox is appearing in the field ; as AI brings new technology into the common fold, people become accustomed to this technology, it stops being considered AI, and newer technology emerges.



“AI needs to be defined...  
Self-learning,  
but is it also self-directed ?”

While the form AI takes on has rapidly changed, the public still remembers the 1950s film-depicted dream of machine intelligence replicating common sense, consciousness, and sentience. Rather than replicating the complexity of human intelligence, current AI technologies aim to analyze more information to predict an outcome in specific domains, rendering parallel and often greater-than-human decision-making processes. For example, the technology behind autonomous vehicles, chatbots, language translation, Netflix film recommendations, and winning a game of Go aim to automate and improve human-level skills in narrow applications and domains.

Given inconsistencies in the understanding of AI, coordinating across actors to devise governance and policy is challenging. Without a common definition of AI, discussing the concerns and implications of this technology is infeasible. Before developing governance strategies, it is necessary to first break down the notion of AI through its key components (data, super cloud-computing, algorithms, hardware, software, etc.), understand the trends in these drivers, and develop specific governance to manage impact of these components on society.

## AI as a socio-technical phenomenon

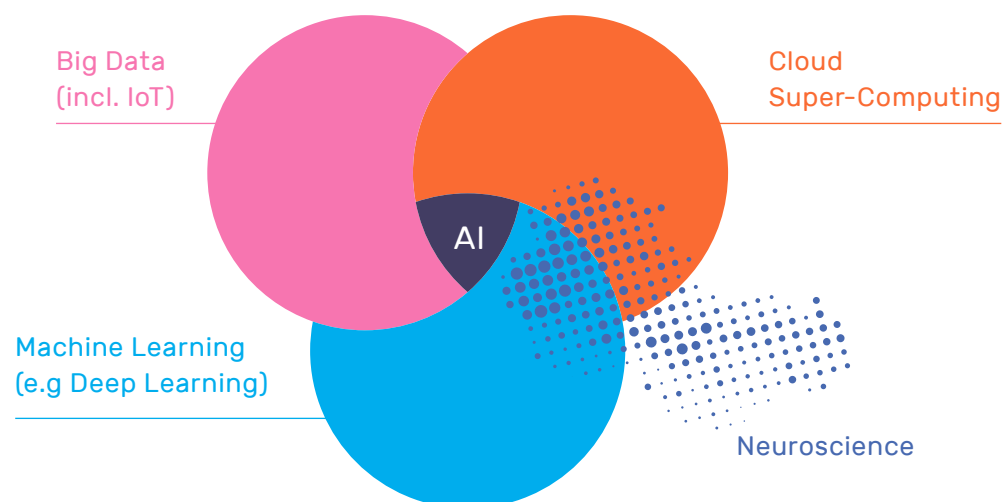
AI reflects interdependence between humans and technology. Recent booms in AI development arise from the convergence of three technological mega-trends: big data, machine learning, and cloud supercomputing. These components do not lie within the vacuum of science and technology. They are defined, created, and implemented by humans. Engineers build datasets, and design, test, and parameterize algorithms, while practitioners interpret output and determine how it is implemented in society. AI should therefore be considered as an interdisciplinary socio-technical phenomenon.



“We are the computers. We are already borged, merged with our technologies. Who wants to be without their smartphone today? Don’t we also want our machines to appreciate and practice values, independent thinking, teamwork, care for others? Maybe they will be better at it than we are.”

Equipped with digital devices, billions of people use, inform, and are affected by AI. Our digital lives fuel machine learning algorithms as we continuously provide vast amounts of data – in the form of social media activity, transactions, and online behaviors. AI is nested in the digital revolution, including social media, attention and platform economy, the internet of things (IoT), and cloud computing, all of which are socio-technical systems based on the pervasive relationship between humans and technology.

AI algorithms have developed alongside advances in computing power. Accessing large volumes of data from digital devices, the Internet, and sensors connected to IoT, machine learning algorithms including deep neural networks could solve even greater challenges, including text and voice translation, computer vision, and image recognition. We therefore define AI as this: *“Big data driven, machine learning algorithm-centric, socio-technical systems powered by supercomputing.”* This definition serves as a lens to assess the dynamics involved in the rise of AI technologies and how to govern them.





## AI beyond human imitation

Several participants raised use-cases for AI beyond mimicking human actions or cognitive functions. Moving away from the notion of AI as replicating humans in form, consciousness, cognition, or capabilities, participants proposed cases for machine learning to complement rather than replicate human activity.

“In my opinion it would make more sense to create AI that is different from humans. That they do not imitate us, but rather that they complete us. This idea is only theoretical, since at a practical level, the vast majority of data collected so far are data related to human behavior.”

The concept of humans and machines that complement each other, rather than compete, was raised as a key goal for how AI is designed. Participants focused on designing AI that completes humans but doesn't necessarily resemble us. In particular, the community was wary of anthropomorphisation of AI and debated how to conceive alternative visions of the technology. The challenge of developing AI that doesn't resemble humans is that the data (the raw material that AI operates on) is based on aggregated human decisions and behavior. Hence, AI algorithms produce outcomes based on analyzing vast amounts of data and spotting patterns within them. So far, we do not have a mechanism to provide alternative data that isn't based on human decisions, preferences, and bias, leading AI to resemble and amplify human decisions.

AI can support human values and goals by approaching problems in new ways. In reinforcement learning, AI technology has optimized scores and reached optimal outcomes in games such as Atari and Go by performing alien moves that humans had not discovered. Early examples of intelligent non-human ways of thinking point to the potential of AI to help solve thus far unresolved global challenges.

Using data from IoT, sensors, digital devices, and other intelligent systems, an AI system may be the most effective agent to optimize complex systems. Early test cases indicate opportunities for machine learning methods to optimize entire systems such as energy, transportation, and public administration. As an example, Google Deepmind has applied reinforcement learning to predict and optimize energy consumption for Google's data centers. In his April 2018 article in Medium, *Artificial Intelligence – The Revolution That Hasn't Happened Yet*, University of California, Berkeley professor Michael I. Jordan argued for the development of intelligent infrastructure. Rather than focusing solely on AI that imitates humans, the technology could take on greater role in coordinating major and complex societal systems.

“By optimizing systems and resources and providing all the non-human-specific work at a very low cost, the AI will make enormous gains in time and money to private and public systems.”

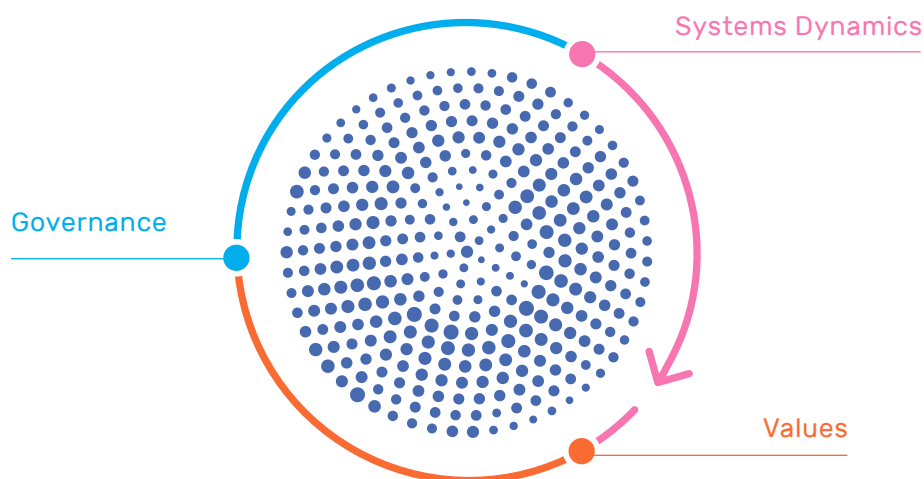
Analyzing innumerable data, variables, and their interactions, AI could become better suited to manage deep complexity as compared to humans. Even a team of humans, each mastering a specific domain, does not have the cognitive capability to have a holistic view over the variables and interactions that make up a complex or chaotic system. With accelerating amounts of digital data and global connectivity, humanity has an opportunity to harness the power of AI to solve our greatest societal challenges.

## AI takes on new political roles: AI-cracy

The digital revolution is opening new pathways for governance. Social media enables new forms of civic participation, lobbying, and mass campaigning. Blockchain technology can be leveraged to securely store public data by non-public sector actors, with a myriad of potential use-cases to improve governance and give greater agency to citizens. AI could also drive new forms of governance through distributed and decentralized digital structures.

“Regulators and policymakers must be knowledgeable ‘citizen champions.’ Then AI could provide information to the current citizen champion to allow for great, efficient and predictive policies over time. If everything is going smoothly, we may not need the citizen champion at all and let the AI do it for us entirely on its own.”

New governance models must also consider evolving citizen values and preferences, and new forms of trust-building. AI technology will become increasingly pervasive in our lives, a shift in our values and norms for how we interact with technology will occur. Changing values in turn impact governance and policy for technology. AI governance frameworks will therefore evolve alongside changing citizen values, which evolve alongside changes in technology.





"We're talking 20-30 years to go through this whole cycle, allowing ourselves time to develop proper AI to do these tasks, gather more data over time, let the AI learn, us gaining trust from the experience and so on, until AI is so involved in good decision making that we let it do it on its own; AI doing a job 10 times better than any human expert at 1,000,000 times the speed. I call that progress."

## 2 Diverging expectations of AI: A call for trust in technology

Across discussion themes – from *AI in the World* to *Forecasting Applications of AI* and *Adapting the Workforce to the Age of AI* – participants' perspectives for the impact of AI on society varied widely, indicating the diversity of global opinion. It is evident there are no concrete foundational moral, ethical, or societal norms that guide a global view on an AI-driven future. People's values and preferences vary, and are informed significantly by country, culture, and other features. Large differences in perspectives could hamper efforts to govern the technology and limit AI innovations. However, deeply-held expectations and values, whether alarmist or aspirational, indicate that citizens should be central to the dialogue on AI governance.



## Forecasting AI : Creative destruction or destructive creation ?



Projections for the impact of AI on society varied between dystopia and Utopia. Participants anticipated the automation of jobs to lead to violence and inequality, or to free us to spend time on meaningful activities. Adoption of AI in the household, industry, and public spaces could lead to atrophying of humans' skills, agency, and control over decisions, or could increase access to basic services and raise standards of living.

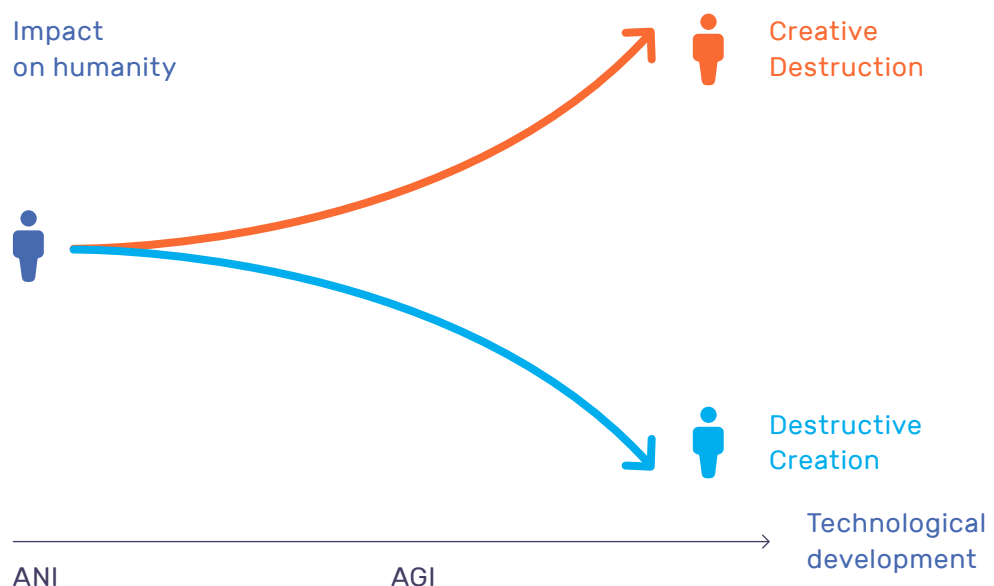
"AI can help end abject poverty."

"We feared losing our souls, our humanity. But it's the opposite !  
Artificial Intelligence has magnified man and offered him the opportunity  
to solve all social and environmental problems!"

"How can regular humans be useful in a world  
where there are others who are 10 times smarter than us ?  
Can plugged and unplugged humans live in equality ?"

These wide expectations highlight the unpredictable and complex nature of the societal impact of AI. As AI evolves, a wide range of potential paths for humanity unfold, with large scale upside benefits and downside risks.

Participants' polarized hopes and fears were particularly visible concerning the impact of AI-enabled automation and unemployment. Some participants expected a post-work Utopia, where people have time to develop their passions, relationships, learning, and exploring rather than working a job to meet material needs. Other participants pointed to large-scale societal disturbances caused by unemployment, heightened inequality, and imbalance of power.



"Workers' rights will change from: *The right to have a decent job and be safe doing it* to *The right to not work just to live and feel safe.*"

"What if machines do the tiring, repetitive work, leaving man the possibility of doing something else? Associations, creation, citizen participation, sports, cultural activities. All this free time allows to do maybe much more beautiful and interesting things?!"

"Automation of jobs will lead to poverty, precariousness, displacement of elderly workers, widespread anger and riot, civil war."

## Differences in values and norms is a challenge to governance



One major challenge in global governance of AI is different risk appetites and interest in AI among populations and governments. Risk appetite is the trade-off between regulation and innovation. It varies across countries, with some players more willing than others to sacrifice citizens' safety to deploy AI. There are also differences at the national level in willingness to trade data privacy for new innovation. For example, 93 percent of Chinese customers are willing to share location data with their car manufacturer, compared to 65 percent of Germans and 72 percent of Americans. China has an ambitious strategy to lead the world in AI development by 2030, and to support this goal, it ensures access to vast stores of citizens' digital data to AI companies.

"Certain values, such as right to privacy, are far more strongly held in Western countries compared to others, where access to services is imperative."

"We need to first have a conversation about what we as humanity value and then collectively figure out what we are willing to compromise together. Only then can we begin to formulate international standards that truly are inclusive and work for everybody."

Moreover, regions with less access to high quality goods and services may be more willing to adopt technological solutions. According to a study by PwC, between 50 percent and 70 percent of Nigerians, South Africans, and Turkish citizens are *willing to have a major surgery conducted by a robot instead of a human*, compared to between 27 percent and 40 percent in the United Kingdom, Germany, Belgium, Sweden, and Netherlands. This variation in the trade-off between values such as individual privacy and data protection compared to better access to essential services enabled through AI presents a major alignment challenge. Governance of AI is critical at the global level but is hard to achieve as countries use AI to straddle growth and development objectives.

## Building trust in AI

As AI becomes increasingly pervasive in our lives, from routine activities to providing outcomes on criminal cases and medical diagnostics, the comfort citizens gain with such technological intrusion is incumbent upon trust. Creating trust in AI is a massive challenge because the complexity in machine learning algorithms is beyond current human comprehension. This complexity results in an acute lack of explainability as to how an AI system reached a conclusion, which its creators are also unable to rectify. Trust and explainability in AI is increasingly important, because its applications can have serious impact on human lives. The lack of trust in AI stems from a range of technological and psychological issues such as complexity in how algorithms operate, potential for bias, and the loss of human agency over technologies. If AI is to benefit people at large, it needs trustability built in and codified, and not simply the perception of trust by the public.



“I would rather talk about real trust and not perceived trust.  
And this is the true challenge. How to build real trust?”

One way to create trust in AI technologies is by not depending on the algorithm to make the ultimate decision, but only use its outcomes for signaling or raising concerns. For example, in healthcare, when doctors do not agree with the indications the AI produces for diagnosis, the issue is raised to a committee that discusses the divergence. The final decision is concluded by domain experts rather than a sole human or AI. Applying AI to confirm human decision is one approach to build trust, and ensuring the algorithm doesn't hold the ultimate decision-making power will alleviate public anxiety over the use of this technology.

In addition, participants suggested that in order to build real trust, collective intelligence in decision-making should be applied. For example, the use of an AI tool could only be possible if it has been certified by a diverse body. This is a similar concept to marketing authorization for medical devices and medicines, ISO standards, or certification of NGOs as being truly of public interest and thus entitled to receive private donations. This is further linked to the notion of setting up an international body, comprised of people from different countries, backgrounds, and expertise, which represent the widest interest possible in society.

To establish trust in models that underpin the AI's output, we also need mechanisms that create transparency and explainability. Users are unaware of how AI has been trained, what datasets are used as inputs, and what biases these could amplify. AI could potentially systemize bias that we experience in society through its use in criminal justice, predictive policing, etc., without accountability. It is important to ensure that AI decisions are explainable



in order for humans to follow the reasoning and logical conclusion and forecast the impact of using AI's output at scale in a forensic manner.

"Personalizing health care needs to analyze each individual's traits. But how do you apply the best treatment? Based on group characteristics. This becomes political. Will an AI base recommendations on races, ages, or ethnicity?"

Our debate community advocated that greater financial resources be deployed toward solving the issue of explainability. Currently, there are no significant incentives for AI producers to resolve the issue and often the argument of intellectual property protection counters moves to create broad understanding of how AI operates. With greater incentives and resources dedicated toward cracking the black box, we are more likely to achieve explainable AI. The role of governments is crucially important in producing economic, regulatory, and governance environments that foster such incentives.

There are several integrated dynamics at play in achieving AI explainability. Public education, controllability, open-sourced technology, and transparency are all important factors. With the convergence of AI, IoT, big data, GPS tracking, and voice and language data, citizens are under constant surveillance and possibly influenced by those who control the technology. Without transparency into what the leading players in AI do with citizens' data and how they use it to influence society, trust in AI technologies is difficult to gain. Some debate participants suggested there should be greater visibility into how AI aggregates data to understand bigger trends. It should be possible to reveal AI source code, along with how it is designed, but not individuals' data. Governments could facilitate this by dedicating resources that detect non-transparent AI systems, similar to the programs for detecting nuclear, and biological and chemical weapons. International treaties could serve as a strong base for supporting this work if the public understands the widespread risk of having unexplainable AI.



"AI has become the essential tool for economic, military, and political competition among humans. Groups developing AI will see transparency as surrendering valuable information to competitors, and there will be intense political resistance to transparency. The only way to counter this resistance will be to convince the public of the danger of social surveillance and control."

"How do we define boundaries between personal and other data with an open-data philosophy? We can't know where our data is ultimately being used and impacting our lives, and thus have less control."

"This video '*AI detectives are cracking open the black box*' of deep learning uses a fascinating example of a video game where the user is a frog. The AI got very good at this game, but it was hard to explain. Then, the AI got to learn how to explain a move based on humans '*explicitly sharin what they were doing*'. What if we could train AI to explain us this way?"



Meanwhile, humans are increasingly aware of their rights when consuming and contributing to AI. Knowledge to understand, critique, and contribute to AI development is key to avoiding technological drifts from our collective values and ensuring that AI is human compatible. Participants argue that this could only be achieved by opening the source code.

"It is essential that the source code of artificial intelligence is open and shared to all. This is the main guarantee to preserve the fundamental freedoms of the human being."



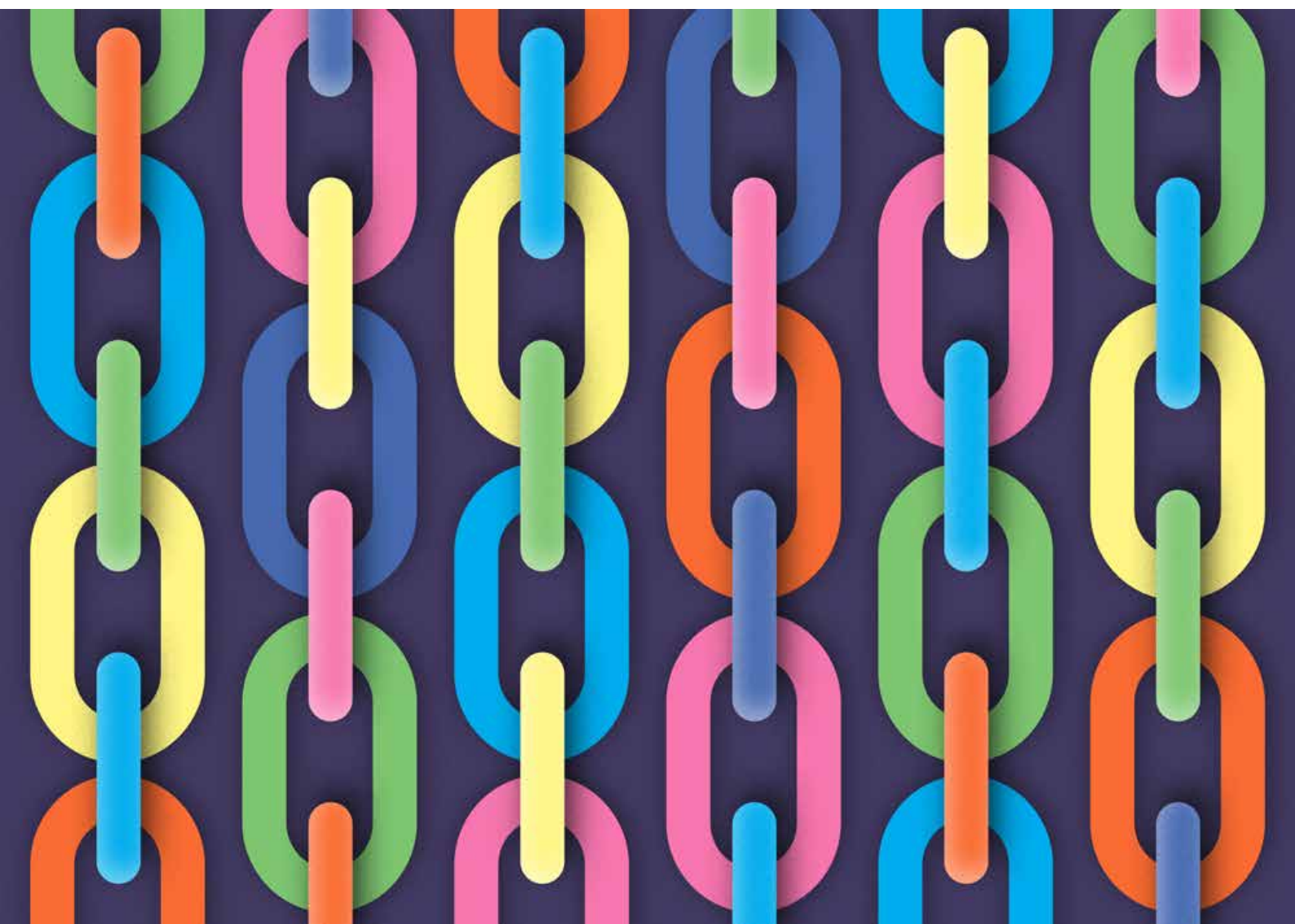
Finally, to foster continued trust in the rapidly advancing AI technologies, timing of introducing technologies is extremely important. If introduction among citizens happens too early, when citizens have not adapted to the use of such technologies, it could induce widespread skepticism, backlash, and possible social unrest. One example is the use of self-driving vehicles, which are already causing media fallout and much discomfort among citizens. The private sector and governments must align on how AI is brought into society and not leapfrog its scaled adoption before preparing the public.

"Introduction of self-driving cars:

- a Industry lobby pushes for proliferation before the infrastructure is ready;
- b Trade unions from transport sector – trucks, public transport and cabs – have huge protests;
- c Many accidents and general safety concerns;
- d Security threats as car systems can be hacked, making it easier to steal cars and potential for new forms of terrorism;
- e Massive protests cripple socio-economic life leading to the government banning autonomous vehicles."

# 3 Applying blockchain for ethical AI and governance

The rapid parallel advancement of AI and blockchain technologies presents cases for these technologies to influence and intertwine with one another. The Global Civic Debate raised opportunities to leverage blockchain technologies to address challenges in AI, such as problems in data security, data access, and algorithmic bias. They also raised the potential to leverage blockchains and distributed ledgers to help govern AI toward safe and ethical development.



Blockchains use distributed and decentralized digital ledgers to record peer-to-peer transactions between parties. This ledger is available and verifiable by all individual entities within the network of peers. Once a block of data has been added to the blockchain, all participants' ledgers are updated with the new information. By leveraging modern cryptography, the blockchain logs are generally indelible and tamper-proof. Its authenticity can be verified by the network of peers, instead of a central authority to process, verify, or store transactions. Open, decentralized and distributed characteristics of public blockchains raise significant potential applications for AI.

A major challenge in AI development is privacy and data integrity for personal data collected and analyzed using machine learning. Since blockchains cryptographically secure transaction data in digital logs, they can be instrumental to help authenticate data and secure it against tampering. For example, health data can be securely transferred from medical offices or biometric wearable mobile applications to a site of analysis pre-approved by the patient.



“Blockchain builds trust in data transfer without passing through a central authority, so we may use its validation system also for AI.”

Blockchain technology could also enable a *data commons* or *data exchange* that provides the infrastructure to securely collect, store, and deliver high-quality data for machine learning, and at the same time provide income streams for those who generate data. With blockchain technology, user data is held in a decentralized storage protocol rather than with a centralized provider. This provides greater security against attacks, manipulation, and removal of the data, and can secure the authenticity of the data once at its source. Citizens may be attracted to offer their personal data in exchange for compensation and transparency into how this is being used, which could improve the quality and labeling of the data. As an example, Berlin-based Ocean Protocol is developing a decentralized data exchange protocol to support generation and access of data for AI for more parties. It uses blockchain technology that allows data to be shared and sold in a safe, secure, and transparent manner.

Challenges presented by using data for AI, such as bias and representation, could also be addressed with blockchain technologies. False or misrepresented data issues could be resolved by the inclusion of a large number of contributors. Having a large and diverse network of peers who review AI development and deployment could ensure greater data representation and curb potential amplification of human bias in outcomes provided by algorithms. Open editing rights for information could lead to greater accuracy and representation, if a higher threshold of participation is achieved. By applying the collective intelligence of a global audience, there is a greater possibility of improving the issues of data currently facing AI development.

Blockchains and distributed ledgers offer a consensus mechanism for a distributed and decentralized network of individuals. Participants questioned whether such a principle could be applied for AI in establishing circles of validators for AI development, such as philosophers, ethicists, sociologists, and economists, among others on a private blockchain. This could avoid having a central authority and enable reaching algorithmic consensus that is driven by digital currency to regulate exchanges or *transactions* in the form of consensus for AI development. Such a decentralized voting protocol could store results publicly or privately in a tamper-proof form.

The peers in such a Blockchain network could review and approve AI systems based on meeting ethical or other defined criteria in a decentralized voting protocol. By selecting a global, diverse, and interdisciplinary community that is knowledgeable on different aspects of AI, consensus on technologies would reflect widespread values rather than those of a few dominant actors. The algorithmic consensus mechanism is automated and transaction data records are immutable and resilient to tampering, improving trust in the process. However, a key tradeoff is the loss of speed and efficiency.

The network of ethicists and key actors could arrive at a consensus about whether the technology aligns with human goals and values before it is deployed. Such a peer group could be incentivized monetarily via a system that awards tokens to those who contribute positively, or via the intrinsic reward of playing a role in validating AI that has ethical and beneficial impact. Such approaches are still early-stage and need further research, but they offer an interesting alternative in governance of AI development.



“Expert miners should not be motivated by the money, but by the ethical role they play by validating an AI process before it deploys its action.”

Participants raised concerns with an accessible open data platform for AI developers. One suggestion to overcome challenges of too much open data was to first restrict the access to highly sensitive personal data as AI is being developed; then engage with a dedicated ethics community for validation of the AI’s use; and finally, use a blockchain-like network of peers for deployment and monitoring.



“AI is such a game changer that the structures should be prepared as far in advance as possible before the rolling out is too comprehensive, when we won’t be able to act as supervision, control and ensure fairness and safety.”

“In short, I am personally betting on structured collective intelligence.”

Blockchain-based mechanisms may also offer applications for the governance of AI. Cryptoeconomics and tokenization introduce mechanisms to incentivize behavior and transactions among players who neither know nor trust one another. In a competitive AI market exhibiting a global race to develop technology as rapidly as possible, where actors distrust each other and are subject to destructive game theory dynamics, these blockchain-based mechanisms may provide helpful solutions to ensure cooperation and rule-following. Such applications have technological and practical limitations at present but may allow for new technology-based governance models that avoid the shortcomings of international agreements and global coordination and cooperation.

## 4 New hopes and Utopian futures

The debate participants hoped that AI will enable significant progress, from scientific discoveries to egalitarian societies. Public aspirations could be leveraged to support governance strategies that shape the development of AI to uphold ethical and safety standards as well as collective values, including equality, dignity, privacy and security.



Hello World !

Delivered



World is typing

## AI can unlock new imaginative, exploratory discoveries and worlds



There is expectation for AI to unlock scientific discoveries, including sea and space exploration or extending life. One pathway is for AI-enabled automation to free resources for humans to undertake such challenges. These hopes mirror projections from popular film and media, including Star Trek, which depicts a futuristic society where material needs are fulfilled, leaving resources for space exploration. It exemplifies the profound impact of media, film and art on shaping societal expectations.

“AI can free up human energy, enabling us to evolve society in new directions or explore places like deep sea and space that we could not due to economic constraints.”



Lofty expectations are in line with new technological advancements. For example, Calico Labs, a subsidiary of Google, uses machine learning and computational science and biomedical sciences to analyze massive biological and medical datasets to gain insights and design interventions for life extension. Technologies that enhance human intellectual and physiological traits may push the brink of humanity, opening the door to the next phase in evolution.

“With the rise of AI, the cost of gene-editing has collapsed. We can sequence the genome very cheaply. This also opens the door to human transformation – can we defeat death?”

## Social progress, including egalitarianism, inclusion, fairness, cooperation and understanding



AI could help close societal divides, boosting equality and inclusion across minorities, the elderly, and citizens in rural regions with limited access to quality goods and services. Automation of human labor could create efficiencies in energy, manufacturing, and supply chains that lower production costs. This implies that more people could access previously costly goods and services, including legal services, medical care, and judicial systems. Machine learning algorithms could enable hyper-tailored prediction at the individual level, supporting access to credit, personalized medicine, and education.

“AI and robotics could be used to serve humanity and provide everyone basic services, products and security universally.”

"I hope AI will be used as a mechanism to create more inclusive societies for individuals of all ages in society."

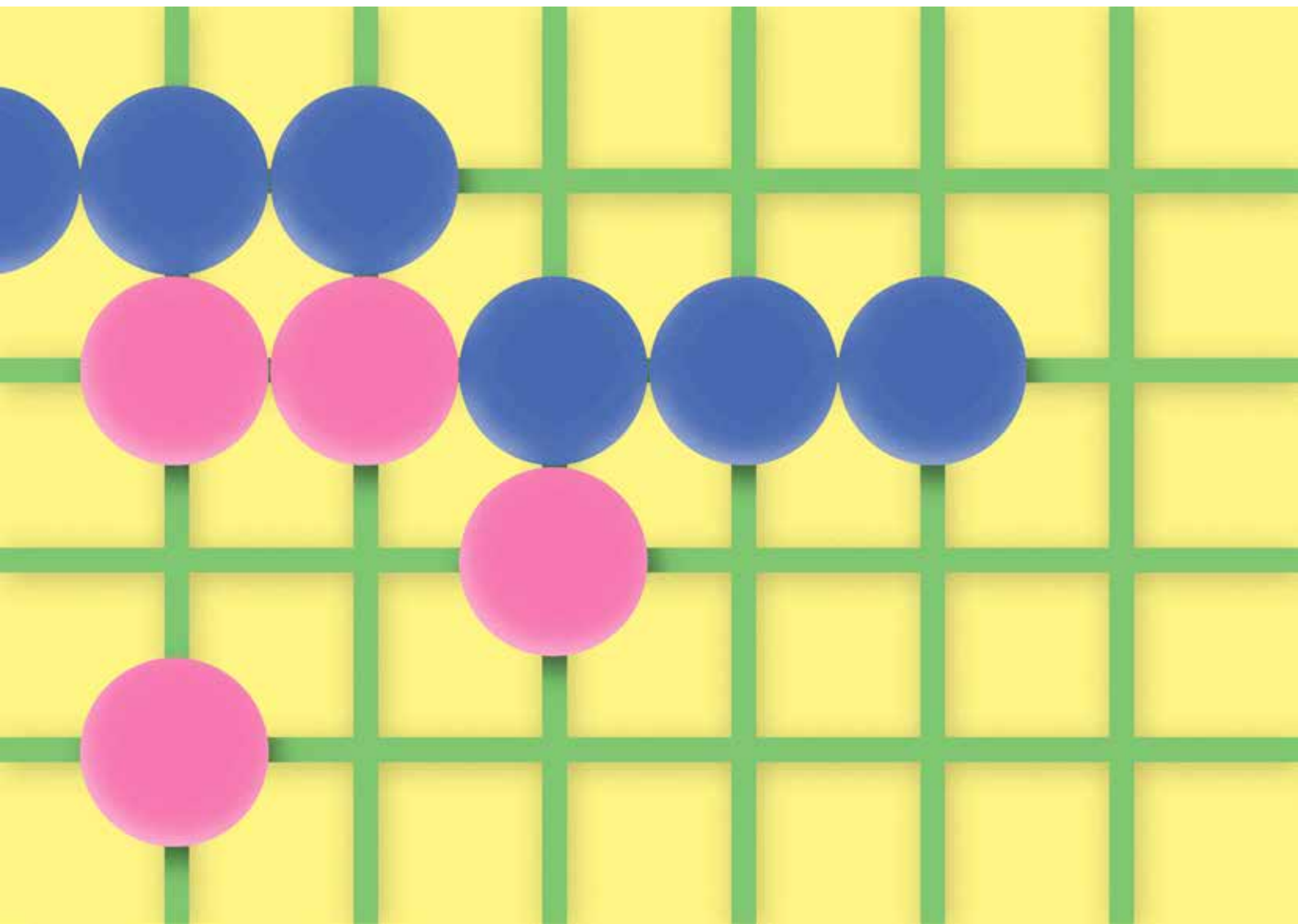


Several participants expressed hope for greater cooperation across society. AI can help solve major global challenges, from climate change, education and healthcare to new sources for economic growth in the face of demographic decline in developed countries. However, unique dynamics in the rise of AI, including a race among states and firms to develop and deploy AI, could hamper progress towards cooperation and equality.

"Are we flourishing as humans in the current society ?  
We need to go from competitive society to cooperative one."

# 5 Smart governance is key to mitigate threats and challenges

The beneficial impact of AI is entangled with downside risks. In healthcare, for example, data privacy and security may be traded off for more accurate diagnostics and personalized treatment. In transportation, autonomous vehicles can reduce accidents and carbon emissions while displacing millions of jobs. Collection of representative, accurate data is often compromised for the sake of more data, leading to algorithms that are biased or unfit for certain populations. In the context of a global race where players aim to develop and deploy AI technologies, governance is necessary to manage the rise of AI to benefit society while mitigating its risks.



## Preserving our collective values of equality and inclusion across society.



Participants foresaw risks in the rise of AI that may threaten our values and societal objectives. In particular, participants feared that ownership and access to data, and AI technology would accrue to a minority elite, which would then gain outsized political, social, and economic power. Policies such as requiring open source code, transparent algorithms, and redistributive economic policies were suggestions to prevent widening inequality, both among and between societies.

“Everyone must have a look, a critical eye and contribute to artificial intelligence to ensure AI respects human rights, democracy and is in the service of universality. You can only do this by opening your source code.”

“New markets will emerge that provide levels of obfuscation to personal data that will lead to more privacy for individuals that can afford it.”

“AI will hit the human economy like a tsunami if it remains in the hands of a few. Only policies can rebalance the wealth produced by AI.”

The concerns reflect a trend of oligopoly market domination by large AI corporations, including American companies Google, Amazon, Facebook, Apple, and Microsoft, and Chinese companies Baidu, Alibaba, Tencent, and Xiaomi. At present, the top AI talent, largest datasets, and greatest stores of computing power are located in innovation hubs centered around these firms. Access to large-scale data poses major barriers to entry for small firms, with network effects driving market concentration. In the U.S. and the E.U., antitrust regulators are considering if large firms such as Google are uncompetitive because of this outsized access to data.

## Competition brings a *race to the bottom* in standards and human values

A global race among firms and states to rapidly develop AI technologies could threaten our values. While competition accelerates AI innovation, without market incentives or global coordination it could also drive a race to the bottom in standards for ethics, safety, and upholding human values. Incentivized to quickly scale, innovate, and offer new products to market, AI players may find it opportunistic to bypass necessary testing procedures and standards to preserve safety and ethical standards.



In the race to develop AI, countries or companies upholding safety and ethical standards may lag. Regulation such as data privacy or safety standards could slow AI innovation. Data protection regulation deprives machine learning algorithms of valuable training data used to improve performance. According to AI safety expert professor Roman Yampolskiy, *“We have a pattern of preferring performance over safety, and that’s what markets usually prefer.”* We therefore must adopt a balanced approach to regulation that mitigates risks from negligence of ethical and safety standards, while fostering positive impact from innovation.

“Is there a world where you can collaborate and have competition, where we still raise the bar and not race to the bottom?”

## Strategies must ensure representative data and accountable scientists

Ensuring representation and inclusion of minority populations in datasets used to train machine learning algorithms is a major challenge. Algorithms trained on unrepresentative data will produce output that does not fit all individuals, leading to suboptimal or even detrimental outcomes when deployed at scale.

Since data is the raw material upon which AI algorithms are layered, using unrepresentative or biased data can produce algorithmic bias that amplifies existing social injustices. Data about humans, collected by humans, reflects inherent biases. With AI increasingly embedded in decision-making processes, such as loan approvals, hiring committees and judicial trials, this bias could be replicated at scale against certain demographics. For example, U.S. judges use a risk assessment tool to predict if a defendant is of higher or lower risk to reoffend, which is considered when setting length of sentence, probation, or parole. However, the test is widely considered to be biased against African-American defendants, who were more likely to be misclassified at a higher risk of violent recidivism whereas their white counterparts were more likely to be misclassified as low risk.

Meanwhile, machine learning algorithms do not produce reasoning behind their predictions. Referred to as *black box*, such algorithms lack explainability and judges cannot provide transparent reasoning behind the predictions. While AI applications aim to accelerate human decision-making processes, debate participants were concerned about sole reliance on opaque, unexplainable, and potentially biased systems.



“The danger is that machine learning is based on currently available – usually biased – data. This does not reduce bias, it systematizes it.”

"It is important to make sure the models underpinning AI decisions are transparent enough to follow their reasoning, or at least to forecast the impact of their decisions at scale in a forensic way to see if it disadvantages certain groups, usually minorities."



Participants suggested that one way to overcome the black box challenge is to call for industry actors and technology developers to be held accountable. Such actors should be engaged and motivated to deeply consider their impact on society; their responsibility does not end in the lab.

"Co-responsibility of the users for the impact of the algorithms, including transparency, pedagogy and open discussions are needed."

## Evolving government and legal institutions for the age of AI

As the AI revolution changes the way humans interact, operate, and manage in society, many of our social institutions must also evolve. To match the pace of technological change, debate participants raised a number of suggestions for transforming governments and legal and social entities to keep abreast.

One way to keep humans, institutions, and AI evolving at a matched rate is by making all information available and monitoring technological development through standards and audits. Participants suggested the establishment of an International Science and Technology organization, which would operate as a collective global body where greater access to AI development information could be made available to a wider number of citizens. Additionally, this transparent monitoring of AI technologies could also help govern and expose potential human rights violations by technology and algorithms.



"On a global basis, we should create ISO standards for AI and human rights and algorithm audits for those standards."

The advantage of such an organization is that it could be set up quickly and could establish new norms through international treaties. However, some participants also contested the notion of an ISO since it may not be an effective way to govern AI technologies. This is because an ISO acts as a certification body when technologies meet the standard and thus, cannot present a barrier to prevent the malicious development or misuse of AI technologies.



"Following game theory dynamics, countries or non-state actors have an incentive to renege and cheat on agreements because of outsized gains from having weapons while others do not and because of concern that others are developing weapons secretly."



Instead,  
participants suggest alternative governance mechanisms, such as:

"Why don't we imagine building up a global network of AI Ashoka-type fellows composed of change makers in AI, with the support of similar mechanisms?"

- 1 Selection of role models;
- 2 Early awareness of AI issues with the very young;
- 3 Teams of teams to conduct large scale  
*AI with positive impact projects.*"



As AI evolves, our approach from a legal perspective will also change. How AI is controlled, protected, and deterred from a legal standpoint remains highly contested. Some participants suggested creating punishable deterrence mechanisms for AI so that these technologies *learn* to act within the bounds of the law. However, such legal implications require significant structural shifts in how our legal systems are set up, the parties of the contracts, and allocation of liability.

"Humanity will get better AI if we treat AI (at least in the beginning) as extensions of individuals or corporations (in the legal sense) that develop them. In the current laws, if an individual uses stolen private information they will be punished. The same should go for AI... At some point in the future that will change as AI obtains rights and is considered a legal entity."

## 6 New threats, fears and risks arise from granting AI algorithms larger roles in society

Participants listed an array of concerns that could arise from AI innovation, including losing agency over data and life choices, human capabilities, disenfranchisement, and coordination failures in critical infrastructure.

These projections reflect underlying anxieties, uncertainty, and in some cases lack of knowledge about AI development. Nevertheless, they are reflective of the unpredictable and fast-paced nature of the technology and its impact. They also indicate the need to build public trust and education.

```
Function(Responsability){  
var Fear =(Risks,Treats);  
for [Build] <society>;  
if choice > (citizen= i+1);  
else capabilities_insert;  
    var output = getValues;  
returnAiDevelopment= ok;
```

## Participants fear insecurity and losing control over major life outcomes



Losing agency over decisions was a poignant anxiety among the debate contributors. As humans increasingly delegate tasks and decision-making authority to AI machines, whether to inform criminal sentencing or credit approvals, AI applications that improve efficiency could also remove humans from the process. Increasing use of machine learning algorithms to inform life-altering outcomes could lead to a fear of loss of control over major life outcomes.

- “Act 1: Google launches a vast development of connected objects (IoT) in individuals. It knows the eating habits and sports of each person.  
Act 2: It resells this data to mutuals and health insurers who can discriminate their customers according to their lifestyle.  
Act 3: Google changes its algorithm to skew the health results on its search engine and create mistrust between patients and doctors.  
Act 4: Google declassifies all health professionals from its search engine and embarks on predictive medicine for all.  
Act 5: No need to consult, Google already knows everything about you.”

## AI brings risks of humans becoming dependent, obsolete, disenfranchised

Participants expressed fear that over-reliance on AI machines could lead to atrophy of our cognitive and physical functions. As society increasingly abdicates tasks to machines, over time our skills and capabilities deteriorate.



At the 2018 World Government Summit in Dubai, professor and co-author of the seminal textbook *Artificial Intelligence: A Modern Approach*, Stuart Russell highlighted a similar risk. “We should be aware of AI enfeebling human society,” he said. “It should be empowering and enabling us to live better lives, but not enveloping us so that we can no longer function as an independent species.”

“Imagine if there’s a robot or a machine that could think for us, remember things for us, make decisions for us. AI could be seen as a tool but could be a downside for us. I would not need to think again !  
I would lose my ability to memorize, to think!”

“We should be careful that the development of AI does not prevent the development of human intelligence.”

Participants were also concerned that asymmetrical access to data would lead to political disenfranchisement. Large masses of citizen data, such as



biometrics, could accrue to an actor or organization. This could then be used to target advertisements and information, including disinformation and fake news, to influence citizens' political, social, and economic choices. For example, historian Noah Yuval Harari said availability of biometric and digital data, computing power, AI, and machine learning, and advances in biology and brain science, now enables the hacking of human beings.

"Step 1: Asymmetrical unlimited access to data (political parties, firms, hackers)  
Step 2: Psychological analysis of data and complete knowledge of preferences  
Step 3: Manipulation of behavior through targeted advertisement and information  
Step 4: People are left with no more choice who to vote for, and politics is predictable  
Step 5: We do not need elections anymore."



Moreover, economic inequality, brought about by technological unemployment, may lessen social and political power. In free-market democracies, employers and jobs provide a platform for engagement and a source of identity and collectivism. Without jobs and employers and lacking material, participants worry about losing power to influence their surroundings.

"Increased economic inequality can leave millions without a voice in the public space."

## Granting AI authority over major infrastructure raises risks of system failures

If machine learning algorithms are given authority over major infrastructure, there is potential for high-stakes errors arising from lack of human-level common sense. Granting authority to coordinate increasingly complex, data-driven, connected systems raises the stakes for safety and control measures.

Machine learning algorithms can consider, analyze, and make predictions for innumerable data points, managing complex systems at levels beyond human capability. However, granting authority to intelligent systems for critical and strategic infrastructure and systems requires robust measures for oversight, testing, and control.



"Since there's no regulation or ethical frameworks, a retaliation by an automated system against a particular threat can be disproportionate; if malicious malware destroys critical infrastructure of another state, the reaction by the defensive autonomous system could be exaggerated and decide to annihilate all the vital infrastructures of the enemy – triggering World War 3."

"How can man step in to override the AI if it becomes too powerful?"



Participants were concerned about disruptions caused by coordination and collusion among algorithms. For example, algorithms trained on the same datasets or set to optimize the same outcomes may drive procyclical events in financial markets. This scenario may be inspired by the 2010 U.S. stock market *flash crash*, which occurred when several frequency trading algorithms attempted to sell at the same time.

"AIs might be trained on similar data and therefore think and act similarly upon market events. This could lead to independent AIs of independent companies buying into the same opportunities and selling at the same time, amplifying the momentum and possibly triggering even more autonomous trading systems to do the same thing. Eventually this would lead to uncontrollably fast bubbles or plunges that deal huge financial damage to all market participants and their environment."

## Embedding human safety in AI



Many participants wanted the development of AI to continue in a manner that captures the benefits and limits the risks of malicious use. Lethal autonomous weapons (LAWs) and malicious use of AI are real threats to humanity, raising pertinent ethical questions. Participants argued that the risks of LAWs must be acknowledged and owned by AI players, across industry and state. This would serve as a foundation for building international cooperation and agreements on prohibiting LAWs. Some comparison can be drawn with the development of the nuclear industry, which failed to own the risks of the technology and ultimately led to industry prohibition on development.

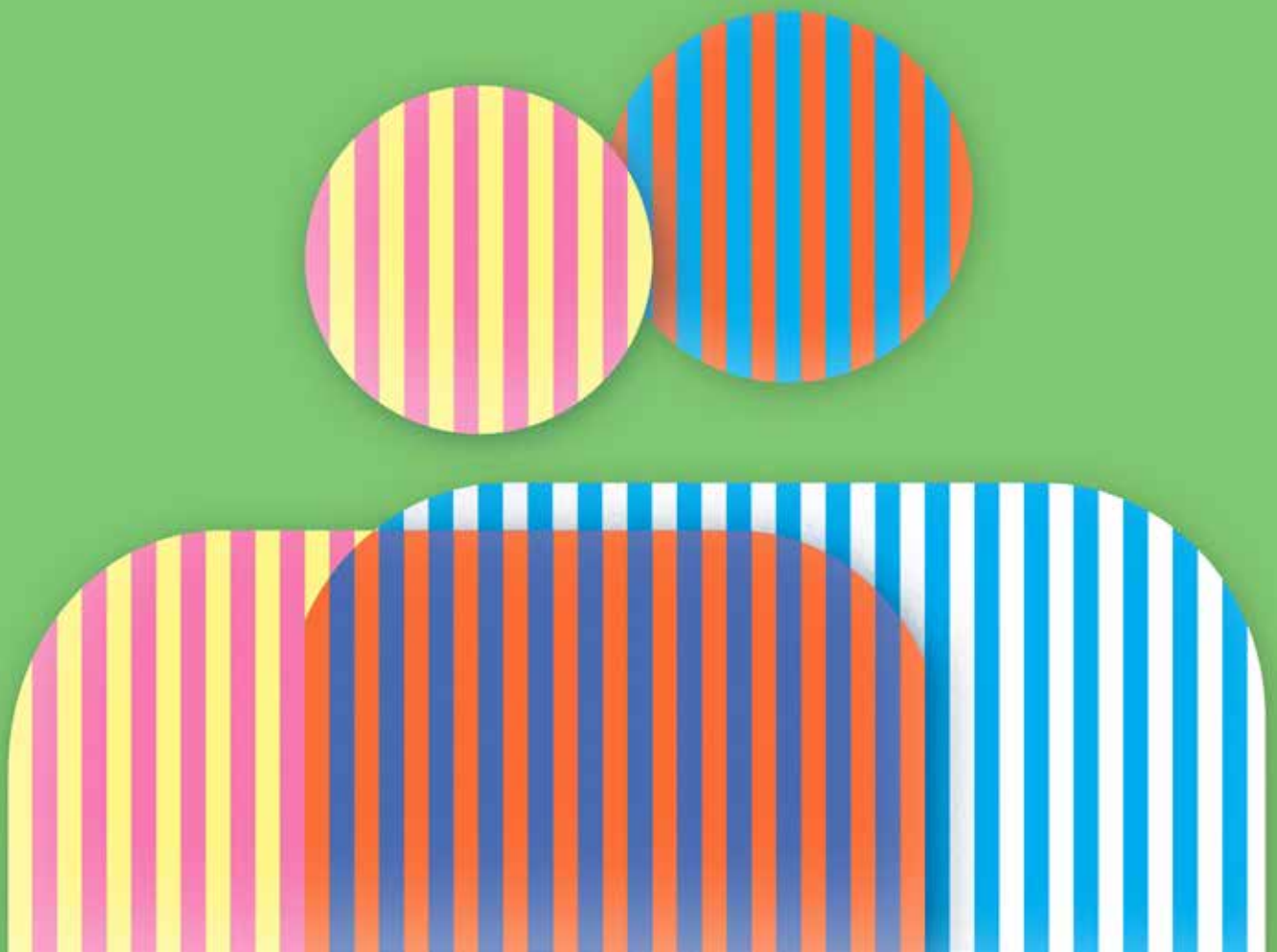
"With the capacity to kill thousands of people very quickly, regulation is crucial. But how can it be achieved and enforced? Is a blanket, universal agreement to ban lethal autonomous weapons preferable, and possible? Meanwhile, robots and military weapons enhanced with AI technologies are already fast under development, for example in Russia's robot army and the autonomous guns at North Korea's DMZ border. How can global cooperation of LAWs be achieved?"

"People need to be made more aware of the human risk LAWs and AI pose and for the AI tech industry to own these. There needs to be a shift in the moral norm, which is what happened in the case of nuclear weapons."

# 7 Evolving identity and roles for humans

What is the role and identity of humans in a post-work world ? How will humans spend their time, self-identify, and what will become their purpose ? What is the role of humans in a society where political, economic, scientific, and other systems are increasingly automated and managed by machines ? How to advancements in bionomics, genomics, and ability to *upgrade* humans re-define humanity ?

Alongside the Internet of Things, connected devices and sensors, synthetic biology, nanotechnology, 3D printing, and other emerging technologies, the AI revolution, nested within the digital revolution, may lead to a paradigm shift in human history and society.



## Identity and role of humans in an AI world

The relationship between humans and the environment in an AI-driven world is set to change, necessitating a re-definition of the role of humans in society. Intelligent Infrastructure may take on new and increasingly present roles; there is potential for AI to take on managing networks or systems, including energy systems, transportation and urban mobility; management of infrastructure; and more. With less agency and control over the systems behind our daily lives, one potential outcome is that the human experience becomes more alienated and enfeebled. The arrival of an Artificial General Intelligence machine that has the potential to surpass human-level intelligence may further paralyze humans' abilities to have control and agency over life outcomes.



"From 2020 the artificial general intelligence will have taken over all the vital sectors of the economy."

AI-enabled automation of human labor and tasks will lead to a re-negotiation of humans' identities and perceived roles in society. Humans may have more disposable time after completing their work and household tasks. Moreover, low-cost production from 3D printing, nanotechnology, optimized supply chains, and energy production enabled by AI and other new technologies may lower costs of goods and services for more people. Need for labor has defined human activity and identity since the advent of agriculture. In the potential scenario of a work-free society, how will humans' identity and roles in society change?



"We live in a society where work is a goal of life in oneself. People tend to introduce themselves, to identify themselves, to seek fulfillment, through their professional activity and the pay that goes with it. This forces us to redefine how we position ourselves in the face of others and the world."

Free from working to meet basic needs, humans may require a fundamental shift in societal norms and values. Today, social, political, and economic models center on work. Urban landscapes constructed around offices and commuter traffic may make way for communities centered on leisure activities. Decline of labor may have implications for gender and group identities. There may be space to reinvent education systems and curricula, which were founded in the 19th and 20th centuries to prepare people for jobs in a post-Industrial Revolution. For example, one potential outcome is AI-enabled personalized education and guidance that supports individuals to explore hobbies and activities that lead to fulfillment.

"Universities still operate like medieval institutions with cumbersome processes that make it very hard and slow to respond to changing demands. We should imagine new institutions and mechanisms to teach the skills and competencies to meet employment needs. Disruption will come from outside, from new models."

"Very few children will turn away if they can fully engage in their passion, which the AI can discover and enrich in lessons."



How will people spend their time if the 40-hour work-week becomes obsolete? Debate participants expected people might spend time exploring hobbies, volunteering, and achieving fulfillment and purpose in new ways. While some social experiments have pointed to negative consequences from joblessness or direct cash transfers, including depression, alcoholism and drug use, it is possible these outcomes may arise from a sense of personal failure in a societal context that praises career success. With the decline in the need and prevalence of work, values and norms may shift toward greater appreciation for learning, socializing, and helping one another.

"We'll promote leisure as a new activity with high personal and social value."

"The state has introduced the universal income and well-being has improved. And what have humans done with all this free time?... Not war, no; not idleness, either... simply: ART!"



With the shifts brought about by AI and other emerging technologies, what will be the role and identity of humans in an AI world? In the book *Superintelligence*, Oxford philosopher Nick Bostrom explored avenues for humans to compete with increasingly intelligent machines, including human enhancement via bionics, selective genetic breeding, pooling humans' consciousness into networks, and more. His outcome is ultimately pessimistic; given the acceleration in computing power and an AGI's theoretical ability for *recursive self-improvement*, or self-upgrading, machines could achieve superintelligence far surpassing humans regardless of attempts at upgrading. In *Life 3.0: Being Human in the Age of Artificial Intelligence*, MIT physicist Max Tegmark predicted several possible outcomes. In all of them, an AI system takes over political control; outcomes range from *benevolent dictator*, *zookeeper for humans*, and annihilation of humans from earth.

"When I overhear people say,  
*We welcome our new computer overlords*, I feel complimented.  
*I like to have my work appreciated*, says the computer."

Humans may remain *guardians of consciousness*, as Tegmark wrote, or merely use AI as a tool for empowerment and self-actualization. The implications are unpredictable, but one takeaway is certain: they will depend on public imaginaries, values, and the governance of AI.

## Re-designing the human-machine relationship

Our community debated several aspects of design and framing of AI technologies that would align with human values, ethics, and diverse preferences. Some participants suggested rules for integrating our values in AI, such as: *"Laws must be simple, simply stated without any ambiguity, timeless and non-contradictory."* However, it was also noted that meeting these criteria make the task of embedding values in AI very difficult.

Many of the participants expressed a desire to create AI that stimulates greater inclusion in society. Participants hoped AI would help include all people and avoid disenfranchised communities remaining on the periphery. Emphasis was placed on alleviating the crisis of integrating the elderly in many Western societies.



"Right now, I see a big crisis in the lack of integration the elderly have in many societies and that simply isn't right. AI, through perhaps physical robot assistance, could help make such groups in society more included in social / leisure / human activities and not feel lonely."

Many participants of the Global Civic Debate contemplated how the human and machine relationship would evolve as our societies, connections, and cognition evolve in the age of AI. In the field of healthcare, some contributors suggested that AI could lead to the control of categorization of people from birth and inhibit life improvement or healthcare choices.



"In the worst-case scenario: In 2050, Social Security could set up a health index to count the DNA capital of citizens at the beginning of their life, then note their habits and mental health. This score would affect the ability to access work, to have children, to access a mortgage, etc. In this scenario where data is predictive, one can imagine the end of the universal health system."

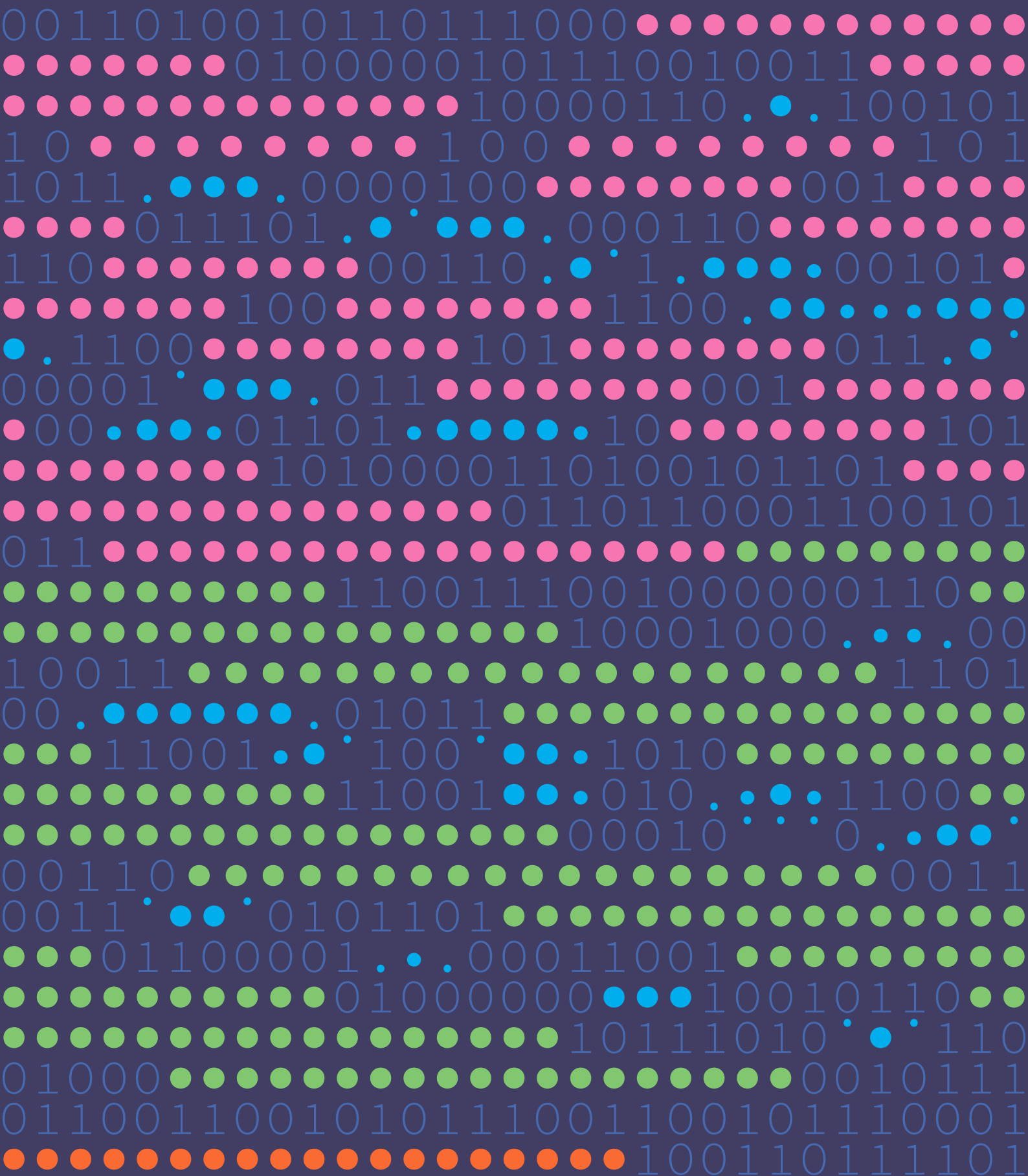
On the upside, participants suggested that AI could enable humans to achieve enhanced capabilities and even immortality. Although today our knowledge of neural operations remains limited, there is a level of convergence between AI and neuroscience, given AI borrows concepts from neural networks in human brains. As the convergence of brain science and knowledge of biochemical processes of the brain rapidly advance, participants expected



AI could help introduce transhumanism, including biological enhancements, bionics and exoskeletons, leading *to upgraded* humans capable of competing with machines. Through brain emulation and using AI, it is conceptually possible to help humans achieve the next phase of evolution.

“With machine learning we will be able to study genes to avoid diseases, to build libraries of correlation, opening pathways to human transformation. We still have 10-15 years as the body remains a black box.”

# Top Recommendations



# Effective decision-intelligence engineering for governance

## 1

Decision intelligence engineering is one mechanism by which governments can achieve AI policies that align with our values. This discipline leverages managerial and behavioral sciences to inform businesses and organizations on how to apply data science, including machine learning and AI. To holistically disentangle the complexity in AI technologies, and adequately address their impact across societies, we need practical tools for policymakers that range across research, convenings, and public education.

The Global Civic Debate led us to support the creation of an Intergovernmental Panel on AI (IPAI), similar to the climate change panel (IPCC) championed by the French government. This organization would serve as a vessel for ideation among governments to create evidence-based policy recommendations for managing the development, deployment, and consumption of AI. The IPAI would embody key characteristics akin to the IPCC: comprise global stakeholders and be supported by multiple governments; backed by the United Nations; focused on evidence-based decision making; and oriented toward policy making. The IPAI would enable global discussion on the dynamics of AI and engineer a pathway toward smart and inclusive policies.

Recommendations from the IPAI could be applied through the AI4SDGs Center, an initiative focused on using AI to monitor, simulate, predict, and act toward the UN Sustainable Development Goals. The center would provide an engine for implementing IPAI's recommendations through real-world projects, embed governance and ethics in AI problem solving, and pursue inclusive development of AI. Through practical application of AI, governments can harness it to help solve administrative and public service delivery problems, which in turn educate government to develop smart AI regulation. Through this *learning by doing* process, government can granularly understand the evolving impact of AI and keep abreast with the pace of change in our society.

# From U.S. versus China to U.S. with China

## 2

AI operates in a first-mover, winner-takes-all paradigm for all players, be they companies or countries. With their diverging approaches, framing, and outlook on AI, China and the U.S. must be brought bring closer together. The AI revolution can follow many trajectories with extreme implications for how we live, our societies, and our futures. The leadership that the U.S. and China have demonstrated in accelerating AI development is resulting in a global race that has profound consequences.

We therefore need to convene technology and business leaders from these superpowers to better understand how AI is developing, how business models are changing with the online platform economy, data economy and attention economy, and the impact of AI on citizens. Such a discourse is essential for creating smart global governance and coordination. The development of AI has significant and concrete risks to global order and international peace. AI is set to trigger a shakeup of geopolitics and agreements between countries, irrevocably altering global power dynamics. Collaboration and international dialogue can help manage and alleviate tensions in the geopolitics of AI while maintaining global order. Consequences of AI superpowers such as the U.S. and China independently pursuing AI development could materialize in restricted international trade and investment, changing dynamics of warfare, growth in AI nationalism, and lack of economic stability, among others.

To initiate this effort, in June 2018, The Future Society co-hosted the inaugural U.S.-China Tech Summit in Silicon Valley in partnership with China's Industry AI Alliance. This day-long forum provided leaders from Baidu, Google, Tencent, Microsoft, and many others a platform for open dialogue on the state of AI industry development in their countries and strategies to coalesce in understanding the AI revolution.

# Global governance institutions

## 3

Governance can help manage the rise of AI among government, industry, academia, and other actors to ensure that equality, inclusion, fairness, safety, and dignity are upheld. A key question posed in the debate was the relevance of global governance. Coordination and governance are needed to balance the complex dynamics involved in AI innovation, including a global race to develop and deploy innovation succeed in competitive markets. Firms are disincentivized from taking time and resources to test and build measures for representative data and safe, secure AI. A *race to the bottom* is developing in terms of ethical and safety standards and human values such as privacy, dignity, agency, and security. Unwilling to lag in pivotal innovation, both states and firms accelerate innovation. Coordination among actors can help raise the bar in standards so that competition yields safe innovation that broadly benefits society.

Governance models range from industry guidelines and soft governance, including technical standards and codes of conduct such as those proposed by IEEE, and hard governance or laws and regulations such as the EU GDPR. Below are ideas proposed by the global community.

### **An ethics committee to create consistent and relevant guidelines**

Participants suggested the creation of a global committee of ethicists to establish guidelines and standards. Today, there are no clear guidelines or review practices to prevent accidental or unintentional harm caused by AI systems. According to ethicist Jake Metcalfe of Data & Society, *“More social scientists are using AI to solve society’s ills, but they don’t have clear ethical guidelines to prevent them from accidentally harming people.”* Currently, U.S. law requires government-funded researchers to obtain approval from an independent ethics committee to conduct research involving humans. However, these guidelines were developed decades ago for interaction with subjects, which is irrelevant to modern data science. The European Commission plans to publish ethical guidelines for AI development this year.

The effort to centralize international standards – initially proposed by Carrick Flynn at EAx Berlin in 2017 – should be guided by three axioms, participants suggested: [1] Funded by member states; [2] Based in a geopolitically neutral state; and [3] Committed to the Common Good Principle, with the flexibility to adapt to changing circumstances. Such an institution would signal commitment by participating states and serve as a focal point for international cooperation. Even if the project failed, it would establish a norm and serve as a cornerstone for future efforts.

“Create a global ethics committee to mobilize an international community of ethicists of human and artificial intelligences.”

“Harness economic fears towards a positive end; it might be a research-first way to force the issue of greater global coordination, including on sensitive strategic matters.”

## A multi-stakeholder approach to governance

A multi-stakeholder approach, with collaboration across government, industry, academia, non-profit, and civil society, increases legitimacy, builds public trust, and bridges the communication gaps across silos. Moreover, broader participation increases capacity by bringing resources and buy-in from relevant stakeholders. For example, independent institutions including academia, nonprofits, and NGOs can inform thinking about fairness, accountability, and transparency. Industry, academia, and other groups can support government and regulation with the implementation of standards and ensure they are applied carefully. NGOs, academia, and others can support assessing if datasets have bias. The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, comprised of more than 100 global thought leaders and experts in artificial intelligence, ethics, and related issues, provides a promising example of a global and multi-stakeholder approach to governance.



“Not just big companies can enforce ethical AI – the effort is more credible with other actors.”

“Have a multi-faceted community (experts, professionals, novices, students, retirees) validate the launch of AI with impact on the human being.”

## An inclusive *collective intelligence* observatory for oversight

A new structure for sustained international, interdisciplinary, and cosmopolitan discussion can promote richer and more complex conversations on key considerations, discard the bifurcation between science and ethics, and reframe the social contract between society and science. It should encompass multiple perspectives from an inclusive international network representing different political cultures and normative frameworks that are not predominantly Western.

The forum would gather information from diverse sources, civil-society groups, and especially the global south; report activities and outcomes from AI developers and researchers; calibrate current methods and approaches

in light of alternative perspectives; and convene periodic assemblies for international discussion.

Conversations would not focus solely on pros vs. cons, risks vs. benefits, acceptability vs. unacceptability, or permissibility vs. impermissibility, but also on how to care for and value human life, individually, societally, and in relation to other forms of life. Discussions should consider whether existing scientific and political institutions are capable of leading the complex deliberations regarding AI, what is needed for genuine societal consensus in terms of representation, and which power asymmetries may shape the future of humanity.



“Establish an International Science and Technology Organization (ISTO) as a globally accessible collective intelligence system. It would make all non-proprietary information on AI available, ensure multiple perspectives, research connected concepts, and expose potential human rights issues through its ability to uncover all AI-related situations.”

## International standards and watchdogs to monitor safe AI development

Participants suggested an International Organization for Standardization (ISO) for AI. At present, the ISO is an international, independent, non-governmental organization forming specialized systems for worldwide standardization. It has published two AI-related standards and is currently developing four others. Participating nations include the United States, the United Kingdom, Russian, and the Republic of Korea.

To be truly collective and representative, participants recommended that such a body should be: [1] Funded by many nations (including the U.S., China, Russia, India, and the E.U.); [2] Located in a geo-politically neutral and/or trusted nation (Montreal and Geneva were mentioned as potentially host cities); and [3] committed to the Common Good Principle with flexible procedure and ability to expand.



“The creation of an agency with a multi-approach (regional, sector-driven and technology-based) would be the most appropriate to ensure cooperation. Criminal deeds will always happen no matter what the structure is in place. But we have to put the safety net as wide as possible, but as accurate as feasible. On the international dimension, as AI should serve the entire humanity (and not just the corporations and technology-minded people and organizations), we should think about how to deploy AI-based technology and solutions under a Creative Commons-type of environment.”

The ISO is an example of soft governance; rather than regulation, it certifies non-binding standards and norms for reliable and safe products and promotes



international uniformity. However, criticisms include that the organization relies too often on general principles, is procedural rather than a tool to minimize the impact of bad management, prioritizes form rather than actual results, and is largely inaccessible except to the business community. Establishing a watchdog for better enforceability of non-binding standards & guidelines is crucial and provides teeth to soft governance measures.

“International watchdogs are vital. Possibly engage the UN, a representative body, on this issue.”

“Create ISO standards for AI and human rights, and algorithm audits for those standards. The ISO is not effective on its own; it’s not a body that prevents processes but rather validates or certifies whether a process respects certain rules. Thus, it will not represent a barrier for the misuse of AI.”

## **Treaties regulating the testing, liability, and use of autonomous drones**

The global community of participants hoped to see the establishment of an international treaty regulating autonomous weapons and drones. A legally binding international treaty to prohibit the development, testing, production, and use of lethal autonomous weapons (LAWs) beyond meaningful human control may have a powerful stigmatizing effect. However, if an international agreement is infeasible, it is also possible that leadership from one or a few leading states could set standards for safe autonomous systems. If a major player takes the lead and prohibits development of lethal autonomous weapons, it is possible that, lacking incentive to develop weapons in this context, other countries may follow suit.



“Banning LAWs is subject to cheating by some countries, but the example of U.S. banning some kinds of research on Human Embryonic Stem Cell worked quite well until today. If countries take the lead in banning those, it may have an international effect.”

“A treaty could create the procedural framework through which nations develop internationally acceptable and safe autonomous air vehicles, or drones, by establishing the necessary tests, data processing capabilities, failsafe mechanisms, and standards of human recognition for AI drones.”

“The establishment of a Treaty on the Use of Artificial Intelligence Drones in Combat is aspirational but critical to securing mutual commitment to ethical AI for future generations. Its design can provide a high moral standard for all nations. It could function like the Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction, which became effective in 1999.”

# Representative and secure data for AI

## 4

To help prevent algorithmic bias, there is a need for data standards, best practices, and collaborative efforts to increase representation and inclusion of broader populations in the datasets used to train machine learning algorithms. Participants suggested several innovative approaches for boosting data representation to ensure that AI systems are generalizable for broader populations.

### **An independent, market-driven, regulated platform for global data exchange**

A data exchange could help address many challenges concerning data for AI. Personal ownership of data that can be sold into an exchange may increase the quantity, quality, and diversity of data. An exchange could also empower citizens to monetize their data, on their own terms, and to know who is using it for which purposes.



“GDPR is a good start, but I prefer the public listing of persons, so that they can take control and profit from their data. This personal ownership of data will increase its quality and more importantly require people to attend to their data appropriate to their disclosure preferences.”

“Bias is implicitly added to results when AI is exposed only to a portion of the data.”

A data exchange could also protect against security and privacy threats by decentralizing collection for example with cryptographically secure blockchains. Moreover, a data exchange could increase access to data for small firms and developers for the training of machine learning algorithms, overcoming a major barrier of entry for new firms.

Some potential characteristics include:

- Independent, fair and inclusive
- Decentralized to increase security, reliability, reduce costs
- Market dynamics self-improve the exchange;
- Inaccurate data and users can be removed
- Prioritize data privacy and security;
- Leverage blockchain technologies to secure data
- Data standards, APIs, and guidelines are standardized across countries and industries

“False data issues are fixed by larger inclusion of contributors. If built properly any system with large enough *flow of data* will filter out false data on its own, as a minor subset of the whole easily identifiable by AI.”

## Blockchain technologies leveraged to secure data and privacy



Blockchain technologies can help secure citizens’ personal data against privacy breaches, cyber-attacks, or attempts to delete or alter the data. Personal data, whether biometrics, location, preferences, or behaviors, sourced from digital devices and the Internet, is unsecured. Blockchain technologies can cryptographically secure and store personal data in a decentralized digital log where it cannot be tampered with, modified, or deleted. For example, medical data can be securely transferred from doctors’ offices or mobile applications to research labs using it for machine learning algorithms for biomedical research or personalized medical diagnostics and treatments.

“Blockchain can build a relationship of trust with data transfer without passing through a central authority, and for validation of personal data used in AI.”

“Validation [using blockchain technology] can occur at each level of application of the AI, for example between an iJINI from Innoplay Lab (South Korea) and the systems deployed in research in defense, nuclear power, genetics, medicine, etc. It can also validate for the identity and geographic origin of data.”

## Education to boost data representation inclusion, and civic engagement



Increasing education about machine learning and AI technologies to a broader population can shift more people into the domain, benefitting the industry and society. Public communication and education about AI technology and its impacts on society help “democratize” AI development and anchor citizens in its benefits and risks. Greater participation can help prevent participants’ fear that the benefits of AI and emerging technologies accrue to a small elite. Diverse perspectives can help remedy potential biases in the design of AI and implementation of AI systems, ensuring they are reflective of the broader public’s values.

“Diversity will be crucial; to ensure that people from different ages, countries, ethnic, and racial backgrounds, gender, education levels, and other demographic groups are involved, aware, and empowered to participate in the new digital economy. For example, AI4All, is a non-profit dedicated to cultivating a diverse group of future AI leaders and promoting a diverse humanistic view of the AI field.”

"We have to prepare education of new technologies to kids. This way, AI will be mastered by the majority and not by a minority elite. If we didn't do that, an elite minority will be able to take the power or to make mistakes for the hearth."

"AI should be open and accessible to all people. Otherwise we cannot ensure that its development will be aligned to human values as it will only reflect the values of certain demographic. This requires deep investment in access and education to the wider population."



There is a need for greater education about AI technology to inform and raise civic engagement. Education about AI technology empowers individuals to manage risks associated with AI for themselves, such as data privacy and security. Education should target and involve all levels of society. Development of curricula or adaptive strategies that consider the variation in access to resources and technological literacy across communities is key.

"A level of digital literacy is important if only to ensure that citizens are aware and informed of the systems and processes behind a new, digital society. Whether or not coding is relevant for jobs, there is value in equipping citizens with the understanding and skills to understand the technologies driving societal changes."

"Promote code learning from an early age under an *ethical* banner."

## AI used to predict its own levels of bias or representation in data



AI and machine learning algorithms can take vast amounts of information and variables into account in making classifications and predictions. One participant suggested that we should leverage this capability to identify if data is adequately representative. Specifically, the participant imagined an AI system that predicts the impact of its output in society. While this may be infeasible today and depends on how humans implement and deploy the technology and its numerous unpredictable consequences, it raises an interesting perspective into an AI system's ability to inform designers of its flaws.

"Imagine a time lapse where an AI system, confronted with possible biased data, could predict the future consequences of data through enhanced computational power that purports false information and observe whether that environment's characteristics are concurrent with our current world and ideals."

# Strategies for mitigating unemployment

## 5

As AI-enabled machines become more capable of performing human job tasks, workforce automation will increase across industries and up the skills ladder. Policymakers must implement strategies to mitigate job losses and their social consequences. Areas of concern include social and redistribution policies, such as universal basic income (UBI) or universal basic services (UBS); labor market policies, such as *flexicurity*; and education and training, including lifelong training that targets adaptability, soft skills, digital literacy, and learnability.

In April 2018, The European Commission outlined a strategy called “*Preparing for socio-economic changes brought about by AI.*” In the report, the commission said it “encourages Member States to modernize their education and training systems and support labor market transitions, building on the European Pillar of Social Rights. The Commission will support business-education partnerships to attract and keep more AI talent in Europe, set up dedicated training schemes with financial support from the European Social Fund, and support digital skills, competencies in science, technology, engineering and mathematics (STEM), entrepreneurship and creativity. Proposals under the EU’s next multiannual financial framework (2021-2027) will include strengthened support for training in advanced digital skills, including AI-specific expertise.

### Cash flow projections to assess feasibility of UBI

Debate participants offered a productive approach to calculate if UBI is feasible. Long-term cash flow projections should include assumptions unique to countries and the rate of technological progress, including decreasing costs of goods and services enabled by advancements in emerging technologies.



“No country has provided a long-term cash flow model based on changing assumptions and parameters yet to assess UBI. The assumptions should vary; for example, the payment of UBI not only depends on the country and region within the country but also the year it begins, since AI can help bring down the cost of transportation, education, medicine, food, construction, and generate new state income, the financial sustainability of UBI is different in 2030 than today.”

## Partnerships between individuals, unions, employers, and government to fund training



As machines become increasingly capable of automating human tasks, citizens will need to up-skill and prepare for new roles on a continuous basis. A suggestion originating from The Millennium Project includes funding training through a collective program involving businesses, government, individuals, and labor unions. In this suggestion, labor unions would collect fees and retrain workers, after which employers may compensate the unions for the training.

“Re-education and training of the workforce can be paid by a combination of business, labor union, government, and the individual. Labor unions can establish collective intelligence systems that list new jobs with training requirements by employers that they expected to offer over the next several years.”

## Local community service movements as an alternative to employment



In a situation of technological unemployment, where professional jobs are scarce, citizens can instead focus on community service, volunteering, interpersonal care, and helping others. In his speech at the World Government Summit in 2018, professor Stuart Russell suggested, *“When machines provide most of the goods and services ... we are going to be directly working with each other to improve each other’s lives. That could be a wonderful future, it’s a very fulfilling life.”* However, Russell also cited the current lack of education and infrastructure to prepare citizens for these new roles. *“We have to dramatically re-tool our education systems and economy to work towards this future,”* he said.

“We need to promote associations and local community service movements as an alternative to standard jobs.”

## ***Adaptability and learnability* education**

To cope with rapid technological change and shifting employment demands, public education should incorporate preparation for citizens to become adaptable and flexible to changing work tasks and requirements. Moreover, education should teach students to *learn to learn*. Learnability and adaptability can help graduates shift jobs as technology changes. In addition, education should be continuous. The learning and working phases will occur simultaneously as adults continue to train and upskill in order to compete with robots and AI for employment.



“Learnability is key; it will be a brand-new world every two years.”

“As AI develops further, automation will also impact up the skills ladder; we are already seeing this with legal work automation and increasing use of AI in medicine. Our workforce needs to be prepared to deal with rapidly changing times and be agile enough to thrive in uncertainty.”

# Tales from imaginaries



Artificial Intelligence is a socio-technical imaginary that is heavily influenced by human hopes, fears, and expectations. Through story-telling and other creative methods, humans are better able to articulate their underlying values regarding AI. *Tales from Imaginaries* is a collective exercise designed to enable the inclusion of broad perspectives from citizens. The projections in our community's stories can heavily inform policymakers, industry, nonprofits, and other stakeholders about the public perception of AI.

The arts and creative industries have a significant role in the way the AI revolution plays out. Hollywood script writers and movie producers play an important part in the way citizens imagine an AI-driven future, which creates a significant influence over their hopes and fears.

Informed by the arts, our collective imaginaries shape our values, preferences, norms, and expectations for technology's impact in society. These in turn influence responsive governance and policy making. It is therefore critical to bring writers, producers, and other artists into the policy making to address the public perception of how AI will impact humanity.

*Imaginaries of AI* was an extremely popular theme among our debate community. Encouraged by this, we organized, along with partner Bluenove, two offline design fiction workshops for collaborative scripting of imaginaries to shape the future. Thus far, our imaginaries project has brought together 500 participants to co-create and produce more than 120 micro-novels and publish 20 scripts in the French magazine *Usbek & Rica*. Here we present a sample of the imaginaries harvested from our community, ranging from fantasies to nightmares. They represent a large collection of submissions that remain to be explored and analyzed in depth.

We also note the acute need for narratives from non-Western cultures. Broadening the conversation can boost participation from other countries in the rise of the AI revolution, making this technology more inclusive and reflective of different cultures. The growth of non-Western narratives broadens participation in the revolution and mitigates the replication of current unjust power structures. It also encourages AI researchers from different backgrounds to balance the demographics in AI research and the types of problems AI technologies solve and produce.

“The conversation of AI Imaginaries could benefit hugely from literary styles such as Afrofuturism, reclamation literature of post-colonial countries, and Muslim-centered futurist narratives presented by, for example, queer artist Zulfikar Ali Bhutto. Afrofuturism is a cultural aesthetic, philosophy of science, and philosophy of history that explores the developing intersection of African / African-diaspora culture with technology. It combines elements of science fiction, historical fiction, fantasy and Afrocentrism with non-Western cosmologies in order to critique and interrogate present-day technological inequities.”

# Great Expectations: Projections of AI

## Know thyself and thy neighbor

Rather than feeding self-isolation in virtual reality and machine companions, the community envisioned that AI could help people better understand their own and others' emotions. For example, chatbots help with mental health and loneliness, while coaching apps can help people identify their hobbies, passions, and career choices.

AI can help with self-awareness and relationships others by revealing covered-up emotions and desires. For example, *A Mirror of the Soul* helped thousands to understand their spouses better and avoid divorce. Participants also envisioned AI helping people understand others through simulations and tailored analyses combining facial, biometric, and verbal responses, leading to *an extreme awareness of others' experiences and connections, like Facebook x 1,000*.

Participants also projected AI helping to unlock and tap into higher levels of human consciousness by revealing meaning in dreams, and the emotions of non-verbal creatures, including our pets and animals.

### | "Living with Your Career Counselor"



March 15, 2050, Advi followed our son days and nights, evaluated him, studied him in all aspects to advise on his career.

- Like any self-respecting father, you want him to go to the stars. Me, I do not know. I do not understand these suggestions: Theology, classical dance or cabinet making ?!
- Hell, I followed all of Advi's suggestions, and not once did she betray me and make a mistake.
- Why don't my parents know me as well as Advi?
- Pierre, we know who will have the last word.
- Yes, you're right darling, the best is to ask him.

<author= "Lili & Keff">

<author= "Laurene">

```
<author= "Miss Sayre">
```

[illegible]

## Live and learn

The community expected augmented and virtual reality to deliver in-person lessons about history, science, literature, and even life. No longer confined to reading from a textbook, people can learn from immersive experiences. Applications for personalized education with AI machines can tailor lessons to be more interesting and meet the learning style of various students. For example, a student can learn about geology with augmented reality of the earth's layers, geometry through interactive experiences with projections of pyramids, or history through projections of historical events. In this case, human teachers may become mere coaches in the learning journey.

### | "Emma de Bovary"



An AI proposes new ways to teach a young teenager with Emma de Bovary, connecting the experience of disconnected holidays boring to the teenager, to de Bovary's boredom: *"After analyzing your problem and emotions, I assembled a selection of cultural masterpieces that tackle the subject from a new angle."*

<author= "Emma">

### | "When in Rome"



**January 15, 2050.** I'm reading a book on the history of the Roman Empire, which I opened with the advice of Zeus, my domestic artificial intelligence. *"Contemporary idleness was already practiced in the year 0: the rich Romans entertained themselves with dances, circus games, and parties that were devoted to the pleasures of the flesh. What if we enjoyed old-fashioned life, like them, rather than stupidly following the daily recommendations of AI?!"* "You have not moved for two hours, you should think soon to walk for a quarter of an hour." "You have already consumed 2,000 calories today: do you really want to eat this cream puff?" I called my friends and the AIs arranged everything; our small group of partygoers experienced a bacchanal session that lasted no less than 42 days without stopping. But gradually, some began to feel disgusted with so many perpetual pleasures. They wanted to leave the experience. After enjoying sex, drugs, and rock 'n roll, they just deflated! I was soaked with gallons of wine, and their attitude exasperated me: I threw myself on one of them, screaming my anger, spat my rage, and squeezing his neck with all my strength. Suddenly, my friend's blue face brought me back to reality: this madness had to stop. I sounded the end of the match. I reauthorized Zeus access to my neural implant.

- Zeus, I need your help... It's a disaster here!
- Well received, I take care of everything: I order cars to accompany your guests home, a robot to remove the waste that litter 90 percent of the apartment according to my visual recognition program; and clean up the place myself.
- Thanks Zeus! `<author= "Sylvie, Hugues, Franck, Perrine">`

## Scientific achievement and Utopian worlds

Our global community envisioned that AI would support scientific improvements that change society. Potentially *our final invention*, participants hoped AI could unlock and support new drug discovery to eradicate cancer and disease. Alongside emerging technologies including 3D printing, AI and robotics can lower the cost of production for houses, food, and basic goods, leading to higher quality of life for more people. With costs of goods declining and AI and robotics automating the tedious human jobs and tasks, people will have time for more meaningful and productive work. With autonomous vehicles and drones and programs to optimize energy usage and efficiency, people will live in safer, cleaner, and healthier environments and communities. Finally, participants also expected AGI and super-intelligent systems to administer specific sectors, including hospitality, journalism, commerce, and healthcare.

### | "Permanent Non-Employment"



January 15, 2050, Year 1 of our new Demo-Créatie, I'm in my living room and enjoying this new state of non-employment. What to do all this time? A balloon hits the window – children play in the square next to my home. They never knew work and will never know it. What strikes me then is that they play: they invent their rules, they create worlds, universes. They just laugh. That's what makes us different from AI. This ability to create, invent. It's the feelings, it's the joy, the carelessness. This idea strikes me, freezes me and makes me suddenly happy. To move on to the next stage, I had to carry this message to the whole world: the new hope of this world would come from our youth, from its candor. For the first time in humanity, we are learning from our children.

`<author= "Raphael, Antoine, Clement">`



```
<author= "Frédéric">
```

[illegible]

```

000000000000
`888'      `8
888          000. .00. .00. .00000. 0000 d8b .00000000 .00000. 000. .00. .00000. 0000 000
88800008 `888P"Y88bP"Y88b d88' `88b `888""8P 888' `88b d88' `88b `888P"Y88b d88' `8"Y8 `88. .8'
888 "      888 888 888 888000888 888 888 888 888000888 888 888 888 `88..8'
888 o 888 888 888 888 .o 888 `88bod8P' 888 .o 888 888 888 .o8 `888'
o88800000od8 o888o o888o o888o `Y8bod8P' d888b `8000000. `Y8bod8P' o888o o888o `Y8bod8P' .8'
                                     d" YD .o..P'
                                     "Y88888P' `Y8P'

```



```
<author= "Frank">
```

## Augmented humanity

Participants came up with stories about a future with human enhancements, where there is a growing difference between augmented humans and non-augmented humans. Augmented humans, perhaps equipped with chips and bionics, will understand more, learn more, and communicate with the outside world. Participants also dreamed of a beneficial Artificial Super Intelligence which enables them to communicate and control outside activities through powerful cognitive capabilities.

| “News by Bots”



Just five years ago, a newspaper 100 percent written by robots became a media empire. *News By Bots* (NBB) appeared in 2045. Led by its mysterious founder EM\_2.5, NBB dominates the media world. Robot or human, theories are numerous as to the founder’s identity, the billionaire entrepreneur himself fueling the rumors. I won the interview of the century! **January 22, 2050**, 14h. I arrive at the headquarters of NBB, a huge technology campus. Surrounded by droids of all kinds, I am guided to a glass elevator climbing to the 71<sup>st</sup> floor. The ascent is stunning, offering a 360-degree view. I contemplate with amazement journalism today: the ballet of the reporter droids transferring to the bots the content of the paper via the centralized cloud, the layout of the newspaper performed automatically live, the TV sets and radio studios filled with animators, columnists and experts but no human. Broadcasted on loop: *“No coffee break, here we carbide the info!” “Mr. EM\_2.5 arrives in a moment,”* warns the voice assistant. A silhouette is then drawn in the beam of light that blinds me. Gradually I distinguish my interlocutor. My eyes then come to rest on his face. Beyond familiar, electrical networks run through his eye, the top of his head and his right ear. I’m facing a man! With a smile on his face, he walks towards me, hands me his hand and says to me in a robotic voice: *“Hello, I’m EM\_2.5, but you probably know me under the name of Elon Musk, no?”*

<author= “Liza”>



It is with great sadness that we accompany the last mortals in their ultimate journey. Everyone follows with great interest this moment of infinite violence. Every person who has decided to live forever has been protected and loved by a humanoid personal assistant. We have let the few last mortals who did not succeed or want to live eternally be free as before. They loved their solitude, their tranquility, their life before. They do not like the idea of progress, empty jobs, enhanced human beings, no longer being the driver of their vehicles, no longer having secrets. They hate the idea of living and being assisted by machines – they want to be autonomous and live with Mother Nature. They eat dead foods and cook dishes. They are easily recognizable because their appearance is not noble, their posture broken and tired, their wrinkled face and features carved by the years. They want to die like their ancestors. The end of elites is the beginning of the divine human, the human god, the immortal human.

<author= “Chay”>

## Our underlying fears about AI

The imaginaries contain windows into our community’s fears and concerns about the rise of AI. One is that humans come to over-rely on machines, which renders them *stupid, useless, or ultimately bored*. Another is that personalized suggestions will funnel people into more extreme *bubbles* among like-minded people, to a greater extent than social media today. Others were concerned that digital and AI systems will know so much data about us, which they can provide to the government or other actors at any moment. Some of these concerns are echoed in the *Key insights* and *Top recommendations* sections of this report, yet they manifest more vividly in written narratives.



January 15, 2050. I was bored to death. My personal AI kept searching for the most wonderful self-fulfilling activities and was presenting them to me each morning at 8 a.m. as an array of exciting experiences to fill the day. I was supposed to be excited. And for some reason, every piece of activity that my dedicated AI identified for me throughout the boundless ocean of opportunities was for me...the ultimate boredom.

<author= “Frank”>



Because you have not eaten in two hours, I would recommend you drink a glass of water. It will significantly reduce hunger, my assistant said. Well, it didn't speak out loud: it was plugged into my internal Wi-Fi and broadcast as a voice in my head. I had chosen a very mechanical tone, so as not to confuse it with my own thoughts, but most people did not bother with such details. In 2050, the assistants' voices in our heads saved us, allowing us to get lost in my activities but still keep track of bodily functions – of humanness. We were freed of doing work but also freed of dealing with our assistants; it was autonomous and trusted. We had forgotten what to eat, when to sleep, how to go from one place to the other. We were happy to delegate some of our consciousness – it couldn't be too harmful, could it?

<author= "Audrey">



Copernicus dislodged you from the center of the cosmos. Darwin denied your biological singularity. Freud showed that you are not in control of your thoughts. And now we robots rob you of the central place you occupied on earth. It's normal for you to have the jitters.

<author= "Anne">

```

      88
     .8'
    .8'
   .8'
  .8'
 .8'
88
0000000000.  o8o      0000      o8o      0000
`888' `Y8b `"'      `888      `888      `888
 888   888 0000 .00000. 888 .00000. .00000000 0000 .00000. .0000. 888
8880000888' `888 d88' `88b 888 d88' `88b 888' `88b `888 d88' `Y8 `P )88b 888
888 `88b 888 888 888 888 888 888 888 888 888 888 .oP"888 888
888 .88P 888 888 888 888 888 888 888 `88bod8P' 888 888 .o8 d8( 888 888
o888bod8P' o888o `Y8bod8P' o888o `Y8bod8P' `8000000. o888o `Y8bod8P' `Y888""8o o888o
                                     d"      YD
                                     "Y88888P'
000. .o o8o      0000      0000      08o      .
`Y8 `"'      `888      `888      `888      .o8
. 0000 000. .00. .00000000 0000 0000 888 .0000. 0000 d8b 0000 .o88800 0000 000
888o. `888 `888P"Y88b 888' `88b `888 `888 888 `P )88b `888""8P `888 888 `88. .8'
"Y88b 888 888 888 888 888 888 888 888 .oP"888 888 888 888 `88..8'
.d8P 888 888 888 `88bod8P' 888 888 888 d8( 888 888 888 888 . `888'
888P' o888o o888o o888o `8000000. `V88V"V8P' o888o `Y888""8o d888b o888o "888" .8'
                                     d"      YD
                                     "Y88888P'
                                     .o. .P'
                                     `Y8P'

```

| "Haves (chips) and the have-nots (chips)"



January 2050, I was reading quietly in my living room, watching my son, a 6-year-old, making his own drone from an open source website he found by himself. My wife, seven months pregnant, can't help herself but thinking about whether or not to implement an electronic chip on our new born. These chips were first introduced by the Chinese by 2040. The Europeans found themselves helpless in front of a new generation, that is smarter and incredibly more.

<author= "Mohamed">

## Our assumptions about AI

The imaginaries and stories uncovered several underlying assumptions that our global community has about AI and its implications. For example, the stories assume that humans will choose to get augmented or not, leaving a polarizing difference between the haves and have nots of chips and enhancements. There are overall sanguine and positive expectations that AI will be benevolent and help humanity to achieve its goals, including better communication, governance, and even to become upgraded humans with better mental and physical health. Rather than mass unemployment, participants' stories are optimistic about human-machine complementary roles in the workplace.

| *"Human-machine complementarity"*



**January 15, 2050.** I'm not tired. I'm a doctor and I'm finishing my second shift of the week. I was 30 in the 2010 years, when the press described a hospital on the brink of collapse, overworked doctors and neglected patients. Hippocrates was my imaginary friend, Hippocrates was my double. In the solitude of my hospitable nights, often I'm prey to indecision. *"Me-Hippocrates"* – together we are infallible. Except that Hippocrates does not need to drink coffee to stay awake. Hippocrates is not hungry. Hippocrates is not whimsical and unstable. Hippocrates is not troubled by the myriad impressions that assail us constantly. Hippocrates does not doubt. Hippocrates is not animated by violent passions of Puccini's operas. Hippocrates has no attraction for adventure. Hippocrates is a robot. Hippocrates contains a memory card and thousands of cables. Hippocrates works on battery. Hippocrates is one of intelligence, a diffuse brain that feeds on data. Hippocrates fascinates me by his inflexible rectitude, the deafening silence of his intelligence made of algorithms. Hippocrates is a chimerical creature: I will never be as intelligent as him. But Hippocrates does not dispossess me, he sublimates me. Because I have what Hippocrates does not have: a soul. Hippocrates and I are the reunification of these two so complementary states that rational intelligence and human intelligence are in a kind of absolute intelligence. It's **January 15, 2050**, I'm sitting peacefully at home reading. And I'm not tired. Hippocrates brought me his intellectual power, I brought to patients my humanity.

<author= "Morgane">

Participants also assumed a level of data privacy and security in their imaginaries. At present, data privacy and security are not major issues for most people. However, it is possible, even likely, that increasing threats of cyberattacks and security breaches will serve as *aha moments* and the public will shift from apathy to action to protect their personal information.

| *"Please show me the vacation video"*



- Okay. Can you tap the thumbnail of the video?  
I need a fingerprint confirmation to have the right to stream a video I found in your family records.
- I press the touch screen, and after a moment the video starts.

<author= "Fabien">



The AI asks, *"Do you want me to put you in touch?"* She hesitates. *"Why not. Log me in."* She takes a closer look at my phone, whose data, like everything else in this conversation, is secure. She smiles.

<author= "Natacha">

# Our methodology and key performance indicators

To foster a dynamic conversation, our civic consultation applied a robust collective intelligence methodology that engaged and motivated our community of experts, citizens, policymakers, and business leaders.

The debate was structured in three phases – discovery, ideation, and exploration – over seven months. The discovery phase provided the community with a survey format interface to gather quick ideas on the themes of AI revolution, AI for the common good, AI's impact on the workforce, and possible futures by 2045. We analyzed the range of responses from this survey and formulated the themes and sub-themes that were explored in detail during the ideation and exploration phases.

In the ideation phase, an argumentative forum was created based on mind maps of the key topics raised during the first phase. This let the community debate each topic and question in detail through discussion threads. In this format, participants could freely respond to any proposal on the platform through counter-arguments and voting on an idea. As we transitioned to the exploration phase, we tackled the most promising ideas emerging from the debate. Since the community and discussions had matured by this phase, we implemented new formats to polarize and synthesize the discussion, including *for and against* formats, alternative suggestions, and open threads.

The debate was managed and fueled by community animators who harvested insights, managed the knowledge shared on the platform, and analyzed the proposals. The community animators ensured the conversation continued to move forward by providing new inputs to trigger the discussion as well as fact-check and structure the discussion.

## 5 languages

English, French, Chinese,  
Japanese, Russian

Engagement & Participation:

## 21 Online & Offline Events

## Global reach:

From China, US, Europe and Russia.

**70 390** Page views

**2 525** Learning hours

**2 200** Participants

**2 074** Votes

**1 291** Written proposals

**702** Writers

Social Media:

## 21 Newsletters

circulated to keep the community  
informed on the latest developments  
in the discussion

**#AlforGood**

and **#ImaginariesofAI**  
reached over **5,5** Million

## 17 Global partners:

**bluenove**  
opening organizations:

**IEEE**  
Advancing Technology  
for Humanity

**WFS** **WORLD  
FUTURE  
SOCIETY**

**future  
of life**  
INSTITUTE

**Usbek & Rica**

**腾讯研究院**  
Tencent  
Research Institute

**nesta**

**CFI**



**CENTRE FOR THE STUDY OF  
EXISTENTIAL RISK**

**CNIL**  
COMMISSION NATIONALE  
INFORMATIQUE & LIBERTÉS



**HARVARD  
Kennedy School**

**LesEchos** **JFAI**

**TheInnovator**

**CNNum**  
Conseil National du Numérique

**ML6**



**UNIVERSITY OF  
CAMBRIDGE**

**Swae**



# Debate health:

## Evaluating the collective intelligence exercise

The Global Civic Debate was our first effort in amassing a global audience to interactively discuss questions of impact and governance for the rise of AI. Over the seven months, this program proved successful in gathering a deeply engaged global community. This was evident in the considerable length and the quality of content posted by participants; many proposals, for instance, were reflexive and tackled the response to a specific question, sharing a detailed perspective. Additionally, several posts embedded videos and external links to news articles or research papers which suggests that participants were highly committed to understanding the subject and furthering collective knowledge by sharing to the community.

Our effort also achieved a broad ensemble of partner organizations from around the world to support the discussion. With 17 global partners, spread across the U.S., Europe, Russia, and Asia, we were able to tap into established communities that are inherently motivated toward the mission of the debate. From workshops held at Tokyo University hosted by the Japanese Society for Artificial Intelligence Ethics Committee, to three global webinars conducted in partnership with the IEEE, our partners actively supported and fueled the debate on AI across their networks.

This awareness-raising effort was further augmented by our two-fold social media and communication strategy. First, we amplified the debate externally by engaging with our community and partners on social media. With campaigns deployed on Twitter, Facebook, LinkedIn, and Instagram, we ensured that the civic consultation could reach all corners of the world. This was mission critical to our goal to foster diversity and inclusion. Second, we adopted an internal community management strategy to nurture the participants on the platform. By regular communication through emails, newsletters, and direct messaging, we cultivated a debate culture to make members of our community feel valuable and part of a global movement. This was validated by the number of thorough and regular contributions made by the participants.

In particular, we were encouraged by the creativity the community poured into the *Imaginations of AI* theme. This story-telling exercise enabled participants to freely envision and write about a positive future for our world with AI. Stories ranged from visions of greater inclusivity, sharing, and equality with the help of AI, and unearthed some of humanity's most profound hopes, fears, and values for this revolution. This creativity transcended the online debate and fueled several design fiction workshops and hackathons around the world.

Although the debate was successful on several fronts and fulfilled our promise to the community, we recognize the challenges and improvements in undertaking a civic consultation. First, this pilot program was an open and inclusive initiative that targeted the public. We faced difficulty motivating the general population to contribute, which is increasingly challenging because of compressed attention capacity with established social media and online forums. Since the debate required contributors to take time to learn about the discussion topics, formulate their own perspective, and then write responses on a separate platform, participation rate was low compared to the awareness we raised through social media.

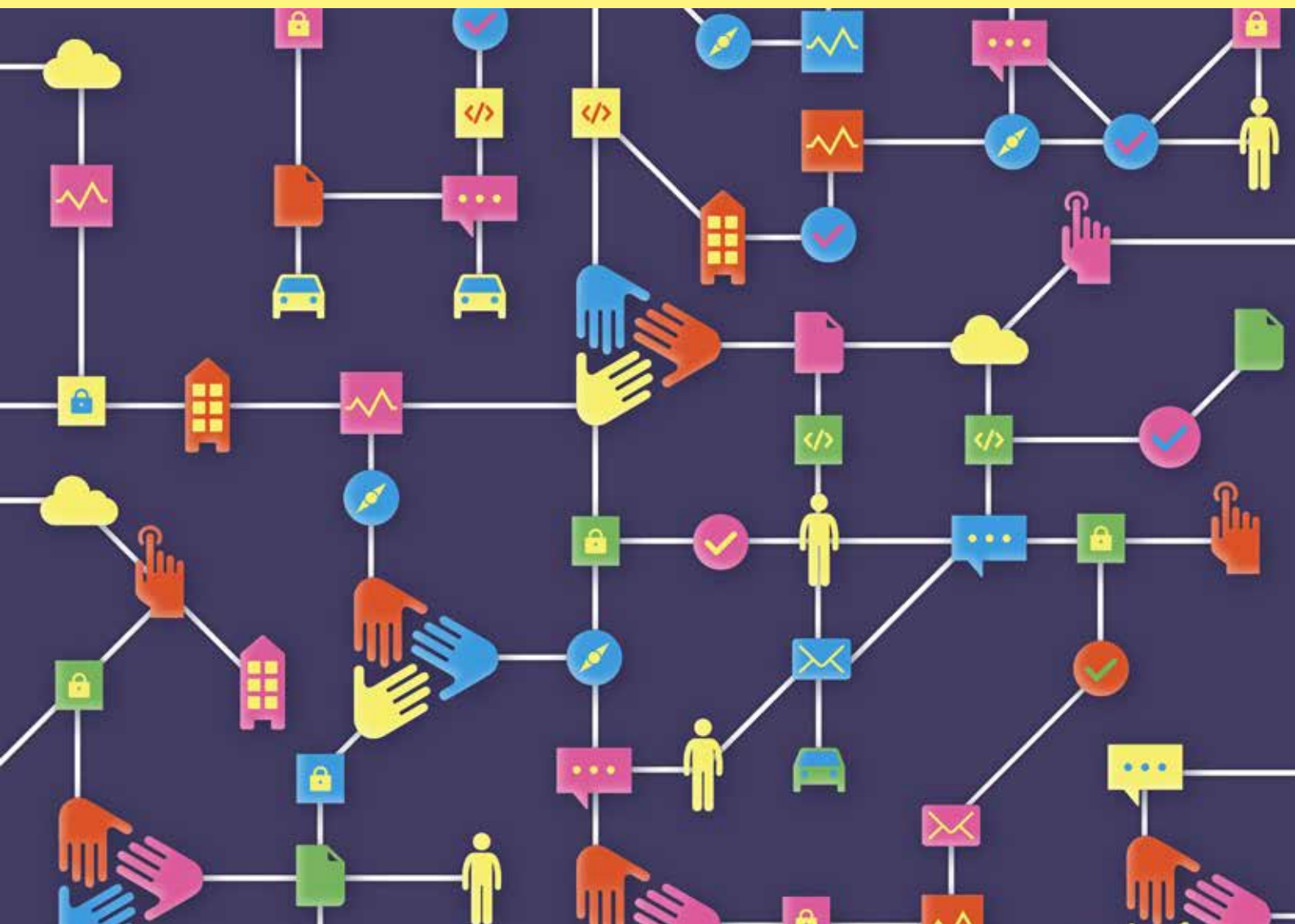
Additionally, we note that AI continues to be perceived as a technical and expert-driven topic which led to potential participants opting to not contribute to the discussion. Though we made significant effort to capture informed citizens through capsule videos, problematizing questions, relatable cases, and visually communicating insights, widespread engagement remained a challenge. This is one of the most significant hurdles of carrying out a public consultation from a grassroots effort.

It appeared *Cybersecurity* and *Frameworks for Governance* were among the less popular themes. This may be because of their technical nature, whereas the community was a mix of experts and lay members. Themes such as *Imaginations of AI* and *AI in the World* garnered significant proposals, likely due to these themes seeming more tangible and triggering creative thought.

Looking forward, we will further our mission of making the discussion on AI more inclusive of citizens by using the insights gathered in the civic debate. Using the power of social media as well as continuing to engage with our global partners, we will disseminate engaging content to further the conversation beyond the public consultation, aiming to shift the public perception of AI from an expert-driven field to a discussion for all of humanity.

# Beyond the Global Civic Debate

The collective intelligence exercise was our initial effort to investigate a thoughtful debate on the rise, dynamics, and consequences of AI. While this report provides a thorough overview of the main insights, discussion points, tensions, and policy recommendations put forward by our community, we humbly recognize that there is much more to scope.



Our promise in this endeavor was to begin a global conversation on AI that is open and inclusive, and we conclude this report with a commitment to continue. With nearly 1,300 detailed propositions on a variety of sub-themes, covering man and machine relationship, AI and the workforce, frameworks for governance, applications of AI, and others, we can continue to mine insights within verticals and across themes. This continuity will manifest in a variety of efforts including serving and building the community within this debate, mining the detailed insights and propositions, and using this initiative to catalyze the global conversation on AI through new research, convenings, and educational programs.

We will amplify the impact of the debate by working alongside experts, thought leaders, industry, and policymakers from around the world to ensure our values, insights, and recommendations inform governance. As we work with policymakers, we will use the community's proposals as key touchpoints to shape the governance frameworks for AI and ensure these achieve a balance of maximizing innovation upside and minimizing downside risks to society.

Complementary to working with policymakers, we will further our efforts at the grassroots level. To ensure the discussion on AI governance continues to evolve in an inclusive manner, we will harness the civic debate insights to create greater awareness of the benefits and key challenges posed by AI. Working with creatives, producers, and visual artists, we aim to use the content on our platform to create images and videos that encapsulate the main insights in an engaging manner. Using our media partnerships and digital and social media channels, we will disseminate this content across platforms to ensure it reaches a wide audience.

As the *Tales from Imaginaries* section depicted, the introduction of creativity in the debate led to an unprecedented response from the community.

With the wide range and high quality of stories exploring possible futures in an AI-driven world, the collective intelligence process provided a robust foundation for building further on these imaginaries. One case is to leverage the stories by working with script writers and producers to create movies on the subjects. As mentioned, we are convinced the arts and creative industry have a critical role to play in shaping the way technologies impact and are perceived by the general population, in turn affecting the way AI is adopted in society. The mass influence held by movie production can be harnessed to springboard positive imaginaries of AI and help the governance process.

We believe that a deep, inter-disciplinary, and multi-stakeholder approach is essential to build smart and wise governance. Therefore, based on our learnings from key portions of debate, we are building spin-off capabilities dedicated to helping large organizations envision, harness, and deploy AI in a way that benefits the company, employees, and their ecosystem.

Finally, we have begun efforts to engineer decision intelligence mechanisms in governance processes through our work across research, convenings, and education on the topic of *Global Governance of AI*. We support the creation of an Intergovernmental Panel on AI (IPAI), adapted from the Intergovernmental Panel on Climate Change. We also support the creation of the AI4SDGs Center, to innovate operational public-private-people partnerships to deploy AI solutions that help achieve the 17 sustainable development goals. These initiatives will operate within the framework of the Dubai Global Governance of AI Roundtable, held under the United Arab Emirates State Minister for AI aegis. The inaugural AI roundtable was held during the 2018 World Government Summit and convened more than 100 AI experts, scientists, practitioners, investors, and regulators from more than 20 countries.

The Global Governance of AI Roundtable community gathered in an intensive day-long collective intelligence forum to explore pathways, challenges, and opportunities associated with the topic. Through our continued work on the roundtable, we aim to foster this evolving community and build on the inaugural convening, as an evolving structure to support the IPAI and the AI4SDGs Center, among other initiatives. Supported by our research and community building efforts, we aim to amplify a holistic and inclusive understanding of policy implications on the global governance of AI through this action-oriented methodology.

As we progress with these projects, we are certain our thinking will evolve and spur further study. Much remains to be harvested, synthesized, and gained from the Global Civic Debate pilot endeavor. We are deeply encouraged by our community to continue and amplify the effort of making a diverse and inclusive conversation about AI that results in technological progress that is beneficial to all.

# THE FUTURE SOCIETY

## Authors:

Nicolas Mialhe  
Arohi Jain  
Caroline Jeanmaire  
Yolanda Lannquist  
Josephine Png

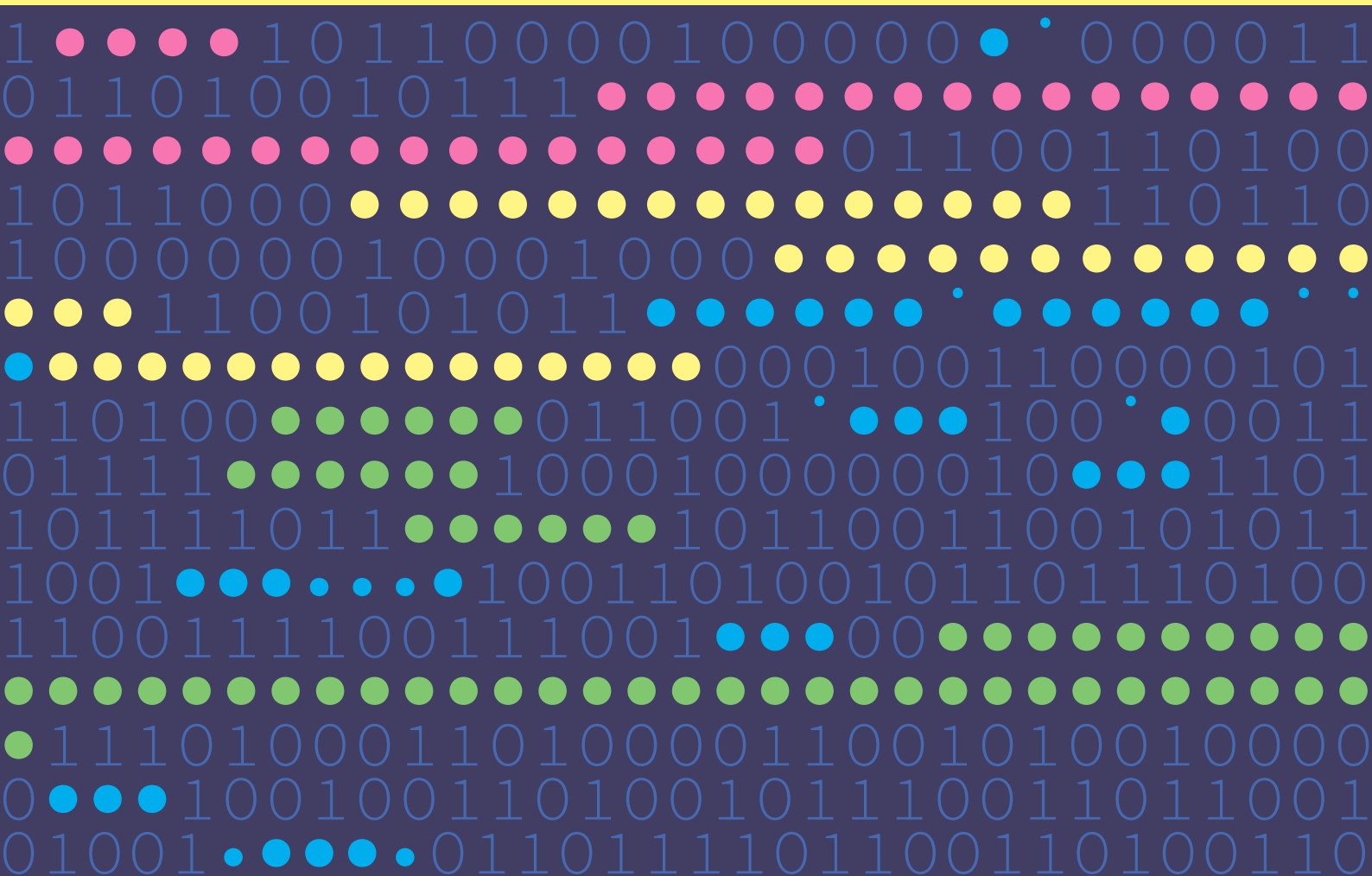
## Acknowledgments:

This report has benefited from the collective effort of many minds and primarily all Civic Debate participants. We would also like to especially thank for their comments and suggestions Cyrus Hodes, Sacha Alanoca, Marie-Therese Png, Marwan Kawadri, Solal Kapelian, Harrold Begg, as well as Frank Escoubes and Bluenove for hosting the Civic Debate online platform.

In addition, The Future Society is grateful to its advisors for their guidance as well as to H5 and Veolia for their partial sponsorship of the project.

We also thank Pablo Hnatow and Mehdi Vilquin for their talented effort in design of the report.

The Future Society © 2018  
[www.thefuturesociety.org](http://www.thefuturesociety.org)



THE AI INITIATIVE