THE FUTURE SOCIETY

Consultation on the White Paper on Artificial Intelligence - A European Approach

Experimentation, testing & audit as a cornerstone for trust and excellence

Contribution by The Future Society, June 2020

The Future Society is a global 501(c)3 nonprofit advancing the responsible adoption of Artificial Intelligence (AI) and other emerging technologies for the benefit of humanity. With a network of policy researchers and practitioners present in the EU (France, Spain, Germany, Belgium, Estonia, Romania), the US and all over the world, we build understanding of AI and its impact, we build bridges between relevant constituents, and we build innovative solutions to help communities and people all over the world enjoy the benefits of AI and avoid its risks.

Overall

The Future Society welcomes the European Commission's White Paper on AI and, more broadly, the European Approach on AI. The European Union is set to become a major leader in the development of this technology worldwide, through its ecosystem of trust and its ecosystem of academic, technical, industrial, and entrepreneurial excellence. We support its approach aiming to balance the imperatives of both innovation and governance in technology. Delivering on these ambitions will require increased technical capacity & ingenuity among enforcement authorities. This contribution recommends actions that could strengthen the case for the European Approach to AI, relying on building the civil service's capabilities for experimentation, testing and audit of AI technologies.

Summary of recommendations

The Future Society recommends the European Commission to establish carefully-designed world reference experimentation, testing & auditing capabilities such as risk assessment frameworks, pre- & post-deployment impact & compliance reviews, benchmarking and calibration protocols, engineering & technology laboratories and test beds. Specifically, we recommend to:

- Review the available evidence for the design of experimentation, testing & auditing policy instruments (such as measurement and EU technology laboratories) and to leverage Member States and other nations' experience.
- Develop and deploy both ex-ante and ex-post compliance mechanisms (such as auditing tools and pre-market testing protocols), and to integrate them into the same governance system for excellence and trust.
- Design experimentation, testing & audit capabilities (such as test beds, benchmarking protocols & risk assessment frameworks) so as to facilitate access and encourage utilization abroad, without lowering the EU's quality standards.
- Design experimentation, testing & audit capabilities (such as test beds, benchmarking protocols & risk assessment frameworks) with and for SMEs, local governments & authorities, start-ups, self-employed individuals, NGOs, researchers, etc., without lowering the EU's quality standards.
- Build agile governance instruments such as experimentation, testing & auditing capabilities designed to adapt to change, using feedback loops such as regular landscape reviews and civic consultations.
- Integrate a programme for Research, Innovation and Competence for trust and excellence within the experimentation, testing & auditing capabilities.

Experimentation, testing & audit as a cornerstone for trust and excellence

The Future Society welcomes the European Commission's emphasis on the development of the EU's auditing capacity and of world reference testing & experimentation centres. If the governance structure is designed properly, we believe these new public capabilities could turn the European Approach into reality, ensuring consumer protection and empowering innovation. Moreover, we believe they could help further additional strategic objectives on the European Commission's agenda.

> Overall recommendation: The Future Society recommends the European Commission to establish carefully designed world reference experimentation, testing & auditing capabilities, such as risk assessment frameworks, pre- and post-deployment impact & compliance reviews, benchmarking and calibration protocols, engineering laboratories and test beds.

We provide more nuance and specific input in the subsections below.

0. Organising experimentation, testing & audit capabilities for success

Societies have had to govern "new" technologies for centuries. While AI is arguably different, a lot can be learned from Member States and foreign nations' past efforts to ensure both trust and excellence. The goals for various technologies have generally been similar, but the strategy, the enforcement mechanisms have varied. The results have ranged from market destruction by overregulation to widespread consumer harm and, sometimes, the successful combination of consumer protection & innovative industry. The European Approach on AI can become a historical success if we pay sufficient attention to the design of the governance instruments used to implement it. In the context of experimentation, testing & audit, these instruments can include EU risk assessment frameworks; pre- & post-deployment impact & compliance reviews; benchmarking and calibration standards and protocols; measurement, engineering & technology laboratories; Member States' testing facilities; and standardized test beds.

Specifically, the objectives of the whole governance system must be clear and regularly re-aligned with society's long-lasting preferences for trust and excellence. Its various governance instruments must have coherent and clear mandates leveraging synergies. The internal and external incentives created by the instruments must be carefully aligned with the overall system's objectives. The organisational capabilities and authority of the instruments must match their mandate (e.g. in order to experiment, test and audit, one must have the competence to establish and administer common

benchmarks and quality standards, which require the competence to establish common metrics and common metrological protocols.)

Many other capabilities beneficial for compliance should be explored -such as third party auditing, AI incidents audit & sharing, audit trails, use of debugging bounties, tools for red teaming.[1] We can and should learn from best practices and governance structure in existing institutions within Member States or abroad, such as the National Institute of Standards and Technology in the U.S. In addition, specialists in the field of Public Administration should be consulted as to how to design effective and novel governance instruments that can enforce both trust and excellence in a technology such as AI.

In brief, we hope the great motivation for the European Approach's ambitions will be accompanied by equivalent efforts to design the evidence-based enforcement mechanisms it deserves.

> Recommendation: The Future Society recommends the European Commission to review the available evidence for the design of experimentation, testing & auditing policy instruments (such as measurement and EU technology laboratories) and to leverage Member States and other nations' experience.

1. Capabilities for ex-ante and ex-post compliance

AI systems are complex products of engineering, with sometimes significant impacts on society. The European Approach should therefore require ex-ante compliance mechanisms, as for all complex and impactful engineering products. In addition, more and more of these systems evolve over their lifetime (learning after deployment). Their impact on society can therefore also evolve in unexpected ways (for example, social media platforms' algorithms have altered the incentives for content producers, resulting in unintentional echo bubbles). By incentivizing investment throughout the lifecycle of the technology, ex-post compliance also promotes excellence in ensuring trust. Ex-post compliance mechanisms are therefore needed to assess that impact.

As both ex-ante and ex-post compliance are needed, the crucial question is how to ensure strategic coordination and consistency between both. As the auditors can learn a lot from the testing & experimentation authorities and vice versa, we expect the EU

[1] For details, see Brundage, M. et al, April 2020, *Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims.* arXiv:2004.07213v2.

civil service to benefit from significant synergies by integrating all compliance-related capabilities into the same governance system.

> Recommendation: The Future Society recommends the European Commission to develop and deploy both ex-ante and ex-post compliance mechanisms (such as auditing tools and pre-market testing protocols), and to integrate them into the same governance system for excellence and trust.

2. Capabilities to promote the European Approach on non-EU technologies

By developing world-class testing facilities with high standards, the European Approach's enforcement mechanisms can become a beacon for American and Chinese developers. The international aspects of the experimentation, testing and auditing facilities could help enforce the technology policy of a geopolitical Commission. However, if it fails to design its governance system for that purpose, it could lead to a fragmentation of the European market away from the global market and disadvantage European players. Through its International Alliance for Human-Centric AI, the European Commission has already highlighted the strategic importance of ASEM[2] & North American technologies' compatibility with European high quality standards with ASEM and North American stakeholders. We welcome the European Commission's efforts in that direction. Given the long-term ambitions of the EU, we however recommend to expand this ambition in terms of breadth and depth.

In terms of breadth, African and Latin American countries are also relevant to the European Commission's objectives and relations. The recent road map for digital cooperation presented by the UN Secretary-General[3] highlights the importance of developing countries in the global debate on AI, and we can expect their roles to grow in multilateral discussions. AI can and should be applied to many Sustainable Development Goals of relevance to developing nations. Recent crises have demonstrated that the quality of life abroad affects European citizens at home. Reconciling the imperative of development with that of the protection of fundamental rights lends itself well to a governance system that seeks to achieve both trust and excellence. Among others, to build upon the work of the EU-AU Digital Economy Task Force, we recommend the European Commission to establish an EU-Africa AI Task

[2] Asia-Europe Meeting, whose membership includes the European Union, the ASEAN secretariat, 30 European countries and 21 Asian countries.

[3] United Nations General Assembly, May 29 2020, *Road map for digital cooperation: implementation of the recommendations of the High-level Panel on Digital Cooperation,* Report of the Secretary-General

Force and Partnership focused on Research & Development, Talent & Skill-building, and Standards.

In terms of depth, thanks to the success of GDPR at exporting itself, the European Commission can be confident in its ability to leverage its critical mass to continue affecting international partners' legislation and practices. The European Approach on AI could experience the same success if its experimentation, testing and auditing facilities are accessible to foreign stakeholders. Moreover, beyond enforcing compliance, the facilities could provide some additional values to stakeholders, such as detailed test results & recommendations for improvements with links to resources, addition to database of tested technologies, etc.

> Recommendation: The Future Society recommends the European Commission to design experimentation, testing & audit capabilities (such as test beds, benchmarking protocols & risk assessment frameworks) so as to facilitate access and encourage utilization abroad, without lowering the EU's quality standards.

3. Capabilities to enforce the European Approach at home

One of the key concerns when establishing new governance instruments is the extent to which "small users" (SMEs, local governments & authorities, start-ups, self-employed individuals, NGOs, researchers, …) are discriminated against. Consistent with its commitment to *subsidiarity* as a key principle to guide the way power is shared between centers and peripheries, the European Approach on AI should be enforced in a way that empowers smaller entities, public and private, unable to shoulder the same compliance costs as multinational giants'.

There are multiple ways to enforce this without affecting the high quality standards expected by EU citizens. As the objective remains compliance for trust and excellence, providing free (or at least largely subsidized) access to testing, auditing & experimentation resources should be considered. Time-efficiency of the testing and auditing protocols must also be considered. Financial mechanisms - such as fee discrimination and cost-sharing with bigger companies - should also be considered.

Moreover, given the ambitions of the European Approach and the nature of the technology to be assessed, new approaches to enforcement that facilitate access could be developed and leveraged (decentralized or distributed testing, cloud-based apps, blockchain, explainable code-reading AI algorithms, etc.) Finally, beyond enforcing compliance, the facilities could provide some additional values to stakeholders, such as

detailed test results & recommendations for improvements with links to resources, addition to database of tested technologies, etc.

> Recommendation: The Future Society recommends the European Commission to design experimentation, testing & audit capabilities (such as test beds, benchmarking protocols & risk assessment frameworks) with and for SMEs, local governments & authorities, start-ups, self-employed individuals, NGOs, researchers, etc., without lowering the EU's quality standards.

4. Consultative and evolutive aspects of the governance instruments

Technology, market applications, socio-economic contexts, and consumer preferences and AI's impact on society evolve over time. What can sound innocuous at a given time (an algorithm that optimizes your feed for content you are predicted to like) can sometimes turn into major harm for society at scale, when other market players have adjusted to the new technology (polarization of society, fake news and echo bubbles). Technologies, markets, socio-economic contexts and consumer preferences sometimes interact in unexpected ways, so that benign technologies combined with each other in novel ways can result in risks greater than expected. For example, the Brexit referendum, the Cambridge Analytica scandal and the COVID-19 pandemic have brought some use of AI in the spotlight. These events might have altered citizens' deep-seated preferences with respect to trustworthy AI.

The EU institutions should hone their ability to study the future of technology, notably through ESPAS, but that would not be enough. To reclaim and maintain its technological leadership, the European Approach to AI's governance system mandate and capabilities should evolve to stay in phase with reality. Indeed, the success of any rule or organisation is not its ability to foresee all the outcomes, but the ability to efficiently adapt in the face of new realities. To do so, experimenting, testing & auditing authorities should institutionalized feedback loops from "reality" on the ground to their protocols. This requires great agility, but can be facilitated if built in ex ante.

Regarding the evolution of the technology and market applications, we could envision an annual report summarizing the state of the technology, leveraging patents and product solutions data. For the quantitative aspects of this research, it would require, among others, metrics and indexing that are comparable over time when it comes to e.g. analytical and compute power. Regarding the evolution of consumer preferences, ensuring periodic, effective and efficient consultations with EU citizens would help shed

more light on the evolution of their deep-seated preferences. It would also increase the governance system's legitimacy and effectiveness.

> Recommendation: The Future Society recommends the European Commission to build agile governance instruments such as experimentation, testing & auditing capabilities designed to adapt to change, using feedback loops such as regular landscape reviews and civic consultations.

5. Research, Innovation & Capacity-building for testing & auditing capabilities

Effective and efficient enforcement of the European Approach to AI will require Research, Innovation & Capacity-building. Experimentation, testing and auditing capabilities will have to enforce a set of more advanced technical requirements in terms of auditability, safety, corrigibility and explainability of AI systems. Most industries would benefit from advances in these fields, even though it might be too costly for them to research them privately. In addition, enforcement will require development and maintenance of common performance, safety and interpretability metrics. The experimentation, testing & auditing authorities should therefore sponsor Research & Development in fields that could advance their ability to enforce the European Approach on AI, either via prize schemes, academic grants or in collaboration with Horizon Europe.

Beyond research, innovation is likely to facilitate enforcement of these technologies. The industry of Regulatory Technology ("RegTech") or organisational innovations could lead to significant progress in cost-effective enforcement of the European Approach to AI - for example, audit software addons, blockchain-based testing, decentralized assessments, code screening, distributed red-teaming, compliance bounties, etc. The use of regulatory sandboxes to test these novelties' robustness in the field should be considered to ensure the occasional failures have a limited impact on trust and excellence. The experimentation, testing & auditing authorities should therefore sponsor the development of innovations that could advance their ability to enforce the European Approach on AI, via public procurement for innovation, prize schemes or in collaboration with Horizon Europe.

Finally, capacity-building is a key precursor to not only ensure excellence in AI but also the trust that comes with ethical and human-centered AI technologies. Safe & ethical operation of these technologies and assessment of their ex ante and ex post compliance will require significant investments in capacity-building for many stakeholders. To achieve sufficient capacity, several options should be explored. For

example, we should consider educating civil servants in Member States, EU and local governments and authorities on AI & trustworthy AI. In addition, to preserve the balance of power fundamental to the EU and its Member States, capacity building is needed in the executive, legislative and judiciary branches of government at all levels, not only in the executive branch. The European Commission should also promote the establishment of EU-wide mechanisms to ensure that humans required to intervene in AI systems -be they developers, operators, sellers or users- have the appropriate level of competence, knowledge and skills for such interventions to be effective, safe and compliant with the European Approach on AI.

> Recommendation: The Future Society recommends the European Commission to integrate a programme for Research, Innovation and Competence for trust and excellence within the experimentation, testing & auditing capabilities.