

**PURPOSE:**

# Input on the European Commission White Paper “On Artificial Intelligence – A European approach to excellence and trust”

**BY: “WONKS AND TECHIES,”**

a multi-disciplinary group at Stanford University, cooperating on international technology and policy issues, led by Ms. Marietje Schaake

---

**JUNE 15, 2020**

**NOTICE**

*Because Stanford HAI is dedicated to promoting a diverse range of voices, the contents of this document do not necessarily reflect the views of our founders, directors, or associates.*



**PURPOSE: INPUT ON THE EUROPEAN COMMISSION  
WHITE PAPER “ON ARTIFICIAL INTELLIGENCE –  
A EUROPEAN APPROACH TO EXCELLENCE AND TRUST”**

Dear Commissioner Vestager, Margrethe,

I hope you are well despite the many challenges Covid-19 presents to Europeans and people worldwide. It is my pleasure to shift focus to the EU AI Strategy, and to share with you input proposed by a multi-disciplinary group of Stanford students and contributors interested in policy and technology.

We share these suggestions while building on a deep appreciation for the importance of European leadership in implementing a values-based governance model for artificial intelligence. These suggestions also support the ambition to strengthen the public interest and AI that contributes to people’s quality of life. Before you is the result of months-long cooperation between students, from freshmen to PhD’s, from computer scientists to law students, with expertise from Stanford faculty and staff. Although each of us has our own emphases and priorities, we all believe in this collaborative effort and in the well-researched proposals in the attached input.

We are available for follow up and happy to assist where suitable going forward.

Sincerely<sup>1</sup>,

Marietje Schaake, Policy Fellow, Stanford HAI, International Policy Director, Stanford Cyber Policy Center

Ruth Elisabeth Appel	Jennifer King
Dathan M. Duplichen	Cindy Kuang
Lisa Einstein	Heajune Lee
Wren Elhai	Shreya Mantha
Muhammad Dhafer Muhammad Faishal	Vidyangi Patil**
Agata Foryciarz	Gailyn Portelance
Sydney L. Frankenberg	Adriana Stephan
Toni Friedman	Alex Tamkin
Zoe Huczok	Alessandro Vecchiato
Kyra Jasper	Eva Zhang
Danielle Jablanski*	Jason Zhao

\*Editor, \*\*Non-Stanford Contributor

---

<sup>1</sup> All of the input, ideas, and recommendations included in this paper are the reflection of the contributing authors’ own research and analysis, and are not meant to reflect the views of any institution or organization.



# Table of Contents

<b>1. Executive Summary</b>	<b>4</b>
<b>2. On Trust, Transparency, and Accountability</b>	<b>9</b>
2.1 Risk Management for Trust and Accountability	9
2.2 Regulation and Compliance Mechanisms	11
2.3 Considerations for Liability and Harm	12
2.4 New Governance Bodies and Responsibilities	13
<b>3. AI Impacts for Production, Skills, and the Labor Markets</b>	<b>15</b>
3.1 Digitization, Access, and Innovation	15
3.2 Jobs and Sectors Adopting Increased Automation	16
3.3 Skills Required for AI Expansion	18
<b>4. Unknown Risks for Widespread AI Adoption in the EU</b>	<b>19</b>
4.1 Unmarrying Military AI and the Broader Economic Market	19
4.2 Preparing Society for Pitfalls of Automation Bias	20
4.3 Testing and Operating in Multi-Systems Environments	21
<b>5. Conclusion</b>	<b>23</b>



# 1. Executive Summary

In our response to the White Paper “On Artificial Intelligence – A European approach to excellence and trust,” we sought to reflect on overarching themes and gaps illuminated by the white paper, to bring attention to potential second and third order effects and guide policymakers toward concrete steps to take in the months and years ahead. As a multi-disciplinary group of academics, military members, technical and policy wonks, we had a diverse group of expertise contributing to each of the sections below. We begin with a

focus on risk and governance for AI in a broad sense, pivot to the known unknowns related to jobs and the AI-enabled economy, and end by outlining a few items categorized as unknown unknowns for the future of AI in the EU. Below is a table preceding our analysis, highlighting the specific recommendations as they appear throughout each section. We hope these reflections and recommendations help to bolster the EU’s initiatives on artificial intelligence.

## ON TRUST, TRANSPARENCY, AND ACCOUNTABILITY

High Level Recommendations:	Page
The determination to classify a system as “high risk” requires a greater distinction between its various potential harms, along a spectrum ranging from high to low risk, and from material to immaterial harm. Regardless of the standard for explainability, auditing should play a role in determining degrees of transparency and ensuring the accountability of systems.	9
As the EU continues developing regulations for automated decision-making systems and other applications of machine learning, checkpoints for citizens’ data and privacy must be built into the process of research, design, and production. These checkpoints could be part of an audit process, and need to be clearly defined to ensure appropriate risk management mechanisms are institutionalized to mitigate harm and enhance trust and accountability.	10
Much like rules governing transparency, standards of “adequate and accessible redress” will be necessary where applications of AI pose high risk to safety. “Ex ante” regulation of a high risk system could require a system to meet certification requirements prior to its implementation in a given use case.	10
New technologies and the datasets that are used should be certified regularly before they can proceed in development and deployment.	10



To assist regulators in determining appropriate accountability mechanisms, an “AI Accountability Act” can set up compliance mechanisms with various laws that enable EU inspections or investigations of AI. Such a law would serve to both cement risk management procedures and incentivize compliance with mandated requirements for establishing safety, trust, transparency, and accountability in deployed AI systems.	11
Portions of these requirements can be addressed simultaneously through the widespread adoption of detailed “Bias Impact Statements” for vendors and potentially future operators of AI systems.	11
In order to establish liability for AI systems which may cause undue harm, the appropriate definition of “damage” should consider inclusion of reference to dignitary rights, invoking Article 7 of the Charter of Fundamental Rights of the European Union. The EU should further incorporate a statement on the need to respect relevant human rights in the context of product liability.	13
The European Commission as well as governments of EU Member States should first conduct stakeholder engagement exercises and additional public consultations.	13
We recommend the creation of an inter-European Parliamentary Committee on AI. While the committee would deal with implementation activities within and across member states, an AI Oversight Board could be created to supervise implementation guidelines.	13
We propose the establishment of an Advisory Committee on Public Sector AI Use to guide actions to responsibly adopt AI in the EU public sector. This advisory committee should also develop a route to advise and share information with the aforementioned inter-European Parliamentary Committee on AI to aid development and implementation principles for the public sector use of AI systems.	13

## **AI IMPACTS FOR PRODUCTION, SKILLS, AND LABOR**

<b>High Level Recommendations:</b>	<b>Page</b>
One effort that the EU can establish is the development of a policy strategy to foster the digitization of production more comprehensively, focusing on sectors that at this moment are lagging in data, computing power, and digitization.	16



**PURPOSE: INPUT ON THE EUROPEAN COMMISSION  
WHITE PAPER “ON ARTIFICIAL INTELLIGENCE –  
A EUROPEAN APPROACH TO EXCELLENCE AND TRUST”**

<p>It is crucial that the EU improves its investment efforts on AI through public and private initiatives. These initiatives must be focused on inclusive innovation and fostering ideas that improve the lives of the many and not the few. For this goal, the EU could institute specialized grant competitions that incentivize and reward inclusive and equitable technological development.</p>	16
<p>In order to protect the rights of workers on the job, it is necessary that the European Union develop a system of regulation of the application of AI technologies for employment and management decisions.</p>	17
<p>It is important to remember that there are many jobs that are impossible to automate and that will be in higher demand in the future, such as jobs that require personal communication, empathy, and creativity. Within the framework of the EU White Paper, the European Commission should include a specific goal to bolster the development of such skills and expand education programs that focus on these skills.</p>	17
<p>The European Union should also consider investment in AI projects aimed at extending the working life of the older population and workers with disabilities.</p>	18
<p>The European Union should seek to develop and foster AI technologies that improve the production capacity not just in terms of efficiency, but also in terms of the quality of the production process and end products, incorporating new labor not typical for certain sectors once certain physical demands have been automated.</p>	18
<p>The expansion of AI job opportunities will, therefore, require increased investment in educational programs in universities and education centers, and online, especially for retaining midcareer workers. The training will require newly adapted skills requisite to work in AI, both in the production and operation of AI systems. This would imply a concerted investment from university consortiums in the European Union European Union to expand graduate programs in fields related to AI, not only by replicating successful initiatives but by also expanding current European research grants to attract students who might be otherwise enticed to study outside the EU.</p>	18
<p>To speed up the adoption of AI technologies by subject matter experts, the EU could develop a series of on-the-job training programs that facilitate the acquisition of AI skills for their workers.</p>	18



## UNKNOWN RISKS FOR WIDESPREAD AI ADOPTION IN THE EU

High Level Recommendations:	Page
<p>It would benefit the EU to consider military AI applications in its broader plans, to prevent and avoid fragmentation in the internal market. The EU now has a unique opportunity to steer the burgeoning field of tech policy compliance, emerging as a parallel to international trade policy, by incorporating global companies with global stakeholders to advance both market economies and regulation authorities on the use of AI.</p>	19
<p>The Revised Coordinated Plan for AI should expand on requirement type C – ‘information provision’ – to include training on automation bias, dual-use and malicious applications of AI.</p>	21
<p>The Revised Coordinated Plan for AI should consider expanding on this leadership to develop widespread and publicly available training on basic nomenclature, development plans, assumptions, and risks for a general audience. The plan should consider dialogues which in turn bring technologists out of research labs and into communities.</p>	21
<p>While systems on systems is clearly a technical and meta problem, the Revised Coordinated Plan for AI should provide a step by step approach to testing systems which will inevitably operate in the same environment. This area of study would be a boon for the member states’ Digital Innovation Hubs and could appoint sector-specific research to each and entice talent to procure, train, and test contrasting models in a sector-specific environment.</p>	22
<p>There is a need for tiered testing phases to include testing systems on systems given that mistakes and biases can result from AI learning in operation. AI systems may also learn from or contrast with other AI systems. This requirement could be developed into its own conformity assessment cycle prior to deployment, meeting requirements for addressees in a more complex testing environment and raising the bar for both security by design and risk prevention.</p>	22
<p>It is necessary to revalidate the European Union’s ability to interoperate digitally in order to execute combined military tests and operations. Asynchronous capabilities currently exist among member states; therefore, a standard for the use of AI in military operations must be established to ensure that new developments do not inhibit interoperability.</p>	22



**PURPOSE: INPUT ON THE EUROPEAN COMMISSION  
WHITE PAPER “ON ARTIFICIAL INTELLIGENCE –  
A EUROPEAN APPROACH TO EXCELLENCE AND TRUST”**

In the case of AI development for military weapons systems, a clear delineation should be drawn between offensive and defensive systems. The EU should conduct a study to determine the areas where increased automation would benefit or weaken infrastructure.	22
Where offensive cyber-enabled capabilities can be enabled through AI, human in the loop decision responsibility must be enshrined. In order to maintain control over military AI tools it is recommended that the EU nest this capability and outline systems intent with an international body, such as the Tallinn Manual. It is further recommended that AI-enabled capabilities remain egalitarian and distributed within the EU.	22

## **CONCLUSION**

We want to see the EU promote an equitable distribution of AI research, development and deployment. We encourage initiatives to increase public awareness, training, and literacy in response to advancements in AI, and suggest the creation of new occupations in the data-driven future. These recommendations can be coordinated and operationalized throughout the EU, made up of distinguished interdisciplinary experts, to tackle the implementation of dynamic policies as they relate to the development and trade of AI hardware and software, cooperation, and the capacity for change. We submit these recommendations for your consideration, and look forward to the European Commission’s comments.	23
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----





## 2. On Trust, Transparency, and Accountability of AI Systems

The white paper’s human-rights based approach to regulating Artificial Intelligence is pioneering. The ethical guidelines mentioned, produced by the High-Level Expert Group on AI, emphasize that transparency and accountability of AI systems must adhere to the notion of “trustworthy AI”. Ensuring trustworthiness for AI will require decisions made by such systems be explainable “in a manner adapted to the stakeholder concerned,” and that mechanisms be put in place to ensure “adequate and accessible redress.”<sup>2</sup>

### 2.1 RISK MANAGEMENT FOR TRUST AND ACCOUNTABILITY

Before considering intended outcomes, there should be a systematic assessment of each data set utilized in AI systems to help mitigate against potential risks of AI developments. Some data sets required for training or testing systems could be considered too sensitive to deploy in comparison to the benefits promised by systems. For example, data used in an algorithm to determine individual and aggregate shopping preferences might be determined to be “low risk” because the potential negative impacts or harms from this information can be mitigated – i.e. personal identifiable information (PII) and privacy can be removed/anonymized. Conversely, sensitive data related to weapons systems for national defense, and the potential for manipulation or

accidents in deployment must be assessed with greater scrutiny. Thus, risk must be assessed in the contexts of each use case, taking into account the consequences of new datasets, other AI systems, and operational environments.

The determination to classify a system as “high risk” requires a greater distinction between its various potential harms, along a spectrum ranging from high to low risk, and from material to immaterial harm.<sup>3</sup> This tiered approach would allow for classification of risks associated with violations of, for example, the right to life, to be treated separately from those associated with violations of the right to privacy, violations to freedom of expression, human dignity and nondiscrimination.

Moving beyond data assessments, transparency in AI decision-making involves a full account of a system’s inputs, outputs, and the factors that led to its decision(s).<sup>4</sup> The EU’s conceptualization of trustworthy AI emphasizes the notion that systems be “explainable” to some degree. This may involve the ability to understand which data inputs have the most impact on an outcome, or whether a specific factor had an outsized effect on an outcome.<sup>5</sup> An interdisciplinary team of scholars at Harvard University recommends that AI systems be explainable a “proportion of the time.”<sup>6</sup> In lieu of an explanation for a system’s specific decision or output, a better understanding of the underlying technology itself should be a bare minimum requirement.<sup>7</sup>

<sup>2</sup> “Ethics Guidelines for Trustworthy AI (High-Level Expert Group on AI, April 2019), <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.

<sup>3</sup> “White Paper on Artificial Intelligence: A European approach to excellent and trust,” (European Commission, February 2020), 10, [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020\\_en.pdf](https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf).

<sup>4</sup> Doshi-Velez, Finale, and Mason Kortz, “Accountability of AI Under the Law: The Role of Explanation” (Berkman Klein Center Working Group on Explanation and the Law, 2017), 4.

<sup>5</sup> Doshi-Velez and Kortz, “Accountability of AI Under the Law,” 7.

<sup>6</sup> Doshi-Velez and Kortz, “Accountability of AI Under the Law.”

<sup>7</sup> Engstrom, David, Ho, Daniel E., Sharkey, Catherine M., and Cuéllar, Mariano-Florentino, “Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies,” (Report Submitted to the Administrative Conference of the United States, February 2020), 75.



## PURPOSE: INPUT ON THE EUROPEAN COMMISSION WHITE PAPER “ON ARTIFICIAL INTELLIGENCE – A EUROPEAN APPROACH TO EXCELLENCE AND TRUST”

While the EU’s General Data Protection Regulation (GDPR) currently provides a framework for the “right to explanation,” it is limited mainly to privacy concerns. The Article 29 Data Protection Working Party Guidelines on Automated Individual-Decision-Making and Profiling, however, could serve as a model for expanding the right to explanation more broadly, and particularly when it comes to AI. Regardless of the standard for explainability, auditing should play a role in determining degrees of transparency and ensuring the accountability of systems. Audits are currently the most effective means of detecting discrimination or other harms in AI systems.<sup>8</sup> Audit trail requirements developed in consultation with academia and industry are desirable, particularly for safety-critical applications of AI.<sup>9</sup> Under this framework, national governments would take on the task of coordinating the regular audit of applications of AI themselves and would certify third-party auditors to carry out assessments. Determined high-risk applications would require government certification while low-risk applications and technologies would simply be required to publish a verified third-party audit.

In addition, the white paper aptly recognizes the potential risks of AI as it pertains to fundamental freedoms as well as personal and collective privacy. The EU should consider adaptations to the way it categorizes citizen data, and invests in data privacy regulation for AI systems. As the data produced by citizens becomes increasingly valuable, companies hold a monopoly on the value they can derive from it. States may begin to classify data as a national resource, and perhaps institute further mechanisms such as data trusts or a “data tax” to be paid to the EU to enable equitable distribution between states. The funds collected from this tax could be used to ameliorate the harmful effects of an economic transition spurred by AI, to conduct research, and to support and retrain workforces. As the EU continues developing

regulations for automated decision-making systems and other applications of machine learning, checkpoints for citizens’ data and privacy must be built into the process of research, design, and production. These checkpoints could be part of an audit process, and need to be clearly defined to ensure appropriate risk management mechanisms are institutionalized to mitigate harm and enhance trust and accountability.

Much like rules governing transparency, standards of “adequate and accessible redress” will be necessary where applications of AI pose high risk to safety. “Ex ante” regulation of a high risk system could require a system to meet certification requirements prior to its implementation in a given use case. If the system is assessed as posing a high risk for safety, regulators may establish requirements for increased explanation of the system, testing phases, and/or education for end users, to bolster explainability and reduce safety risks. Alternatively, an assessment of high risk for discrimination may require a comparison with the human equivalent to the system, whereby AI outcomes are measured against human decisions and results.<sup>10</sup>

There should be an independent third-party institution responsible for making risk assessments for safety, trust, transparency and accountability. New technologies and the datasets that are used should be certified regularly before they can proceed in development and deployment. This risk classification system would not relieve companies of prospective consequences related to potential harms, but can be developed to mitigate the most extreme cases of harm. It can also serve to promote and proliferate the idea of privacy by design in engineering.

Ensuring comprehensive accountability for AI depends on understanding which elements of a system can be clearly and

<sup>8</sup> Casey, Bryan, Farhangi, Ashkan, and Vogl, Roland, *Rethinking Explainable Machines: GDPR’s ‘Right to Explanation’ Debate and the Rise of Algorithmic Audits in Enterprise* (Berkeley Technology Law Journal 34, no. 145), 183.

<sup>9</sup> Brundage, Miles, et al., “Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims,” (Cornell University, April 2020), 3.

<sup>10</sup> Engstrom, “Government by Algorithm,” 77.



credibly demonstrated, through what mechanisms, and with what tradeoffs.<sup>11</sup> A broad range of accountability tools are available to regulators today, ranging from Algorithmic Impact Assessments to “Bias and Safety Bounties” that incentivize sharing of unexpected and/or unintended behavior by AI systems.<sup>12</sup> To assist regulators in determining appropriate accountability mechanisms, an “AI Accountability Act” can set up compliance mechanisms with various laws that enable EU inspections or investigations of AI.<sup>13</sup> Such a law would serve to both cement risk management procedures and incentivize compliance with mandated requirements for establishing safety, trust, transparency, and accountability in deployed AI systems.

## 2.2 REGULATION AND COMPLIANCE MECHANISMS

Three levels of analysis requirements outlined in the white paper provide a basis for addressing algorithmic bias and discrimination: training data, keeping of records and data, and information provision. Portions of these requirements can be addressed simultaneously through the widespread adoption of detailed “Bias Impact Statements” for vendors and potentially future operators of AI systems.

### Features of Bias Impact Statements:

- Detailed description of the AI application’s designed use cases and use domain
- Considerations of diversity and equity in the dataset(s), especially for requiring mandatory reporting on data distributions of protected characteristics such as ethnicity, age, and gender
  - A standard list of protected characteristics for specific types of use cases can be established

for AI systems, based on GDPR and other international human rights frameworks

- Results of the AI model from mandated technical metrics used to define and quantitatively measure fairness
  - We recommend establishing a requirement to include multiple metrics to represent as many aspects of equity as possible for determining a fairness measure. Potential metrics could include anti-classification, classification parity, calibration, equality of opportunity, and disparate impact
- Establishing pathways for users to flag and challenge discrimination issues
- Evidence of a notice and commitment to solicit public third-party review
  - We recommend the formation of a group of engineers across Member States that facilitates expert review of such notice and commitment releases

Bias Impact Statements, pioneered by the Brookings Institution and AI Now<sup>14</sup>, include detailed information regarding the datasets used, the intended use case and domains, the process of training, and sample results as evaluated on specific fairness metrics for developing AI systems. As the white paper notes, the High-Level Expert Group has published a set of guidelines and a corresponding “assessment list for practical use by companies.” However, this assessment list is intended to prompt general reflection, and compliance is non-binding. Mechanisms for compliance, such as Bias Impact Statements, should be mandated, especially for designers and vendors whose systems are classified as high risk.

<sup>11</sup> Brundage et al., “Towards Trustworthy AI Development,” 4.

<sup>12</sup> “Brundage et al., “Towards Trustworthy AI Development,” 19.

<sup>13</sup> European Parliament, “Tools for Ensuring Implementation and Application of EU Law and Evaluation of their Effectiveness,” (Directorate-General for Internal Policies, 2013).

<sup>14</sup> Nicol Turner Lee, Paul Resnick, and Genie Barton, “Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms,” *Brookings Institution*, May 22, 2019, <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>.

## 2.3 CONSIDERATIONS FOR LIABILITY AND HARM

The white paper correctly identifies the specific challenges that AI systems pose to the classic legal liability model. Prior to artificial intelligence systems not working as they are expected, or causing harm, it is essential to determine who is responsible, who is liable, and who must be held accountable for their unintended consequences.

In the realm of products and services liability, the following causal chain is often charted:

Figure 1: Products and services liability causal chain<sup>15</sup>



When arbitrating liability in a court of law, the plaintiff must demonstrate that they suffered tangible harm, demonstrable either in monetary value, or bodily injury (3). Secondly, they must show that the harm was caused by a default in the product (2). Finally, it must be proven that the given default in the product was in fact caused by a fault of the producer, via negligence or tortious intent.

Due to the “autonomous” dynamics of machine learning, harm produced by the outputs of an AI system is not always predictable or foreseeable. One approach to liability would be to consider that AI machines have moral agency and can be treated as “artificial persons” and potentially ‘punished’ accordingly. Mostly inspired by theoretical cases involving

autonomous cars or weapons,<sup>16</sup> this lens is not intuitive or suitable to most of the legal community. A more viable option would consist of applying strict liability, imposed on the producer even when no fault is discovered on their part before the harm is caused. Both would require a reporting mechanism that would, if nothing else, allow for valuable lessons to be drawn from the processes leading to the harms.

The strict liability model is employed in the EU Product Liability Directive with respect to products (defined as “movables” in the text.)<sup>17</sup> Extending it to AI-powered software and systems would solve part of the liability conundrum. Because of the “many hands” problem distinctive of AI – i.e., the multiplicity of people and parts involved in creating each system, and the interdependence of systems on other systems – ascribing liability to a single producer will be difficult. As a result, the EP’s Committee on Legal Affairs<sup>18</sup> has proposed a collective responsibility framework in which producers of AI systems pool their resources to compensate plaintiffs. This would need to include third party suppliers of subcomponents. In such cases, producers bear collective liability, without being morally responsible for harming consumers. However, collective liability may be counterproductive for product security. Producers could see the obligation to pay out liability claims as an excuse to forego implementing features of security by design in production. To mitigate this, producers who demonstrate a “good faith effort” in respect to guidelines for a system’s risk assessment could reduce their degree of liability based on a set incentives structure.

The EU Product Liability Directive streamlines liability for the producers of defective products throughout the EU by introducing a system of strict liability, in which the injured party is entitled to compensation if he or she demonstrates

<sup>15</sup> Figure 1 is the author’s representation of this causal chain for products and services liability. This figure was informed by Keating, Gregory, “Strict Liability Wrongs,” in *Philosophical Foundations of the Law of Torts*, edited by John Oberdiek (Oxford University Press: May 2014), 292-310, and “Products Liability,” Cornell Legal Information Institute, accessed May 31, 2020, [https://www.law.cornell.edu/wex/products\\_liability](https://www.law.cornell.edu/wex/products_liability).

<sup>16</sup> Dremljuga, Roman & Kuznetsov, Pavel & Mamychev, Alexey. “Criteria for Recognition of AI as a Legal Person,” *Journal of Politics and Law*, Vol 12, No. 3, (August 18, 2019):10.5539/jpl.v12n3p105.

<sup>17</sup> European Union, “Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products” (July 1985).

<sup>18</sup> Committee on Legal Affairs, “Report with Recommendations to the Commission on Civil Law Rules on Robotics” (Report submitted to the European Parliament, May 2016).



damage, a production defect in the product, and the causal link to the resulting damage or harm. In order to establish liability for AI systems which may cause undue harm, the appropriate definition of “damage” should consider inclusion of reference to dignitary rights, invoking Article 7 of the Charter of Fundamental Rights of the European Union. The EU should further incorporate a statement on the need to respect relevant human rights in the context of product liability. Extension of liability for damage caused by AI-powered software and systems must be coupled with acknowledgement of the broader harms to individuals or societies that can be caused by developing technologies.

Since the harms caused by AI will not always be physical in nature, their harms will not always be clearly demonstrable. For example, being mistakenly identified as an offender by an AI-powered facial recognition technology system may be a costly experience, even if the error is eventually corrected. Law enforcement, healthcare, and other applications of AI systems may cause harm to a person’s reputation or honor, or subject them to “indignities”. In 2014, *Google Spain v. AEPD*<sup>19</sup> ruled that dignitary rights could provide a basis for liability, by upholding the “right to be forgotten” implicit in Article 7 of the Charter of Fundamental Rights of the EU. A similar approach could be included in the European Civil Code when AI harms affect individual dignities or inflict emotional distress.

## **2.4 NEW GOVERNANCE BODIES AND RESPONSIBILITIES**

Considering the different levels of government within the EU, it may be difficult to consolidate the engagement of all relevant stakeholders. To overcome this challenge, the European Commission as well as governments of EU Member States should first conduct stakeholder engagement exercises

and additional public consultations. It will be vital to include the full range of relevant stakeholders, including civil society organizations. While EU laws apply across the Union, localized contexts and implementation patterns for new technologies may slightly vary. In order to encourage streamlining and cooperation, we recommend the creation of an inter-European Parliamentary Committee on AI.

While the committee would deal with implementation activities within and across member states, an AI Oversight Board could be created to supervise implementation guidelines. An existing model for such a board could build upon the European Data Protection Board (EDPB) which oversees the implementation of GDPR, and has outlined guidelines on data privacy, while providing member states flexibility in implementation. The committee, like the EDPB, would be composed of EU representatives and regulators. The implementation mechanisms of GDPR could be built upon to consider enforcement mechanisms for AI regulations. Outlining specific responsibilities between the implementation decisions and enforcement bodies is essential for avoiding duplication of efforts regarding the regulation of AI.

The white paper outlines the Commission’s goal to initiate an open and transparent dialogue to help facilitate “deployment, experimentation, and adoption” of AI by the public sector, as well as an “Adopt AI Program” that will support public procurement of AI systems. Such a dialogue should be a continuous and evidence-based effort that leverages available field expertise. Therefore, we propose the establishment of an Advisory Committee on Public Sector AI Use to guide actions to responsibly adopt AI in the EU public sector. This advisory committee should also develop a route to advise and share information with the aforementioned inter-European Parliamentary Committee on AI to aid development and implementation principles for the public sector use of AI systems.

<sup>19</sup> *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González.*, Judgment of the Court (Grand Chamber), Audencia National, (May 2014).



PURPOSE: INPUT ON THE EUROPEAN COMMISSION  
WHITE PAPER “ON ARTIFICIAL INTELLIGENCE –  
A EUROPEAN APPROACH TO EXCELLENCE AND TRUST”

The Advisory Committee on Public Sector AI use will serve as a platform to leverage multidisciplinary expertise to coordinate policy across EU member states. Members could be recruited experts who are offered grant funding for research in exchange for serving on the committee. Specifically, the committee should have the following responsibilities:

- Assess and issue guidance on which use case for AI, especially in context, should be promoted or regulated in line with the treaties
- Issue coordinated guidance on the procurement of AI technologies in the public sector
- Assess ways the EU and member state structures could be adapted to facilitate novel applications of AI – e.g., reduction of bureaucracy, easier access to data for research purposes, increased funding for AI initiatives, novel crowdsourcing initiatives
- Share best practices and lessons learned in the public sector deployment of AI
- Monitor ongoing use of AI in the public sector and potential issues like use case creep or uses of data that were not intended
- Make recommendations to governments and various agencies on the regulation of AI in their respective fields



## 3. AI Impacts for Production, Skills, and Labor

The EU White Paper on AI highlights the importance of developing an ecosystem of excellence and an ecosystem of trust in the implementation of policies that address the challenges posed by AI. In order to achieve these two important goals, it is crucial for the EU to design policies that limit the threats to job security and displacement, and to enact policies that foster research and development while also allowing a whole of society approach to reaping the benefits of transitioning sectors to these new technologies. The adoption of AI is still at its infancy for many sectors, and while the pace of adoption is likely to rapidly increase, the EU has the opportunity to influence the widespread adoption of AI in the EU by optimizing policies that will provide the groundwork for equitable and diffused growth in the AI-enabled future economy.

### 3.1 DIGITIZATION, ACCESS, AND INNOVATION

A 2018 study from a European consulting group portrayed a strong European AI ecosystem, with over 2,000 startups, hundreds of labs and 3,000+ AI communities in 44 European countries. “Trends in investment flows demonstrate the extent of interdependency within the European ecosystem, as well as its interconnections with the global leaders in AI, namely the United States (US) and China. A coordinated investment, talent and regulatory strategy would strengthen the European AI ecosystem and set Europe on a clear path towards global leadership.”<sup>20</sup> Increased investment, research and

development, and new patents could all stem from increased focus on and partnership with European AI startups. Venture capital (VC) funding is widespread in only eight member states, Denmark, Finland, France, Germany, the Netherlands, Spain, Sweden and the UK, with the average European venture capitalist fund only equaling half of the average American VC fund.<sup>21</sup> The Stanford HAI Index Report<sup>22</sup> further reveals that most EU member states have no reported patents as of 2019, and Germany, considered a leader in AI, currently has half the number of AI patents per capita compared to the United States.

Some estimates suggest that AI “could contribute up to \$15.7 trillion to the global economy in 2030, more than the current output of China and India combined.”<sup>23</sup> Increased productivity and increased consumption will be brought on by the automation of tasks currently performed by humans, and increased productivity of manufacturing and services. These benefits will not apply to all companies and all workers equally, and, in fact, the introduction of AI has been flagged by many as a significant threat to equality overall. Even for businesses, the barrier to entry for adoption of AI is not equal for companies from a resource outlook, with adoption of AI ultimately feasible at different points of production.

A prerequisite for AI is access to computing power for research, development, testing and implementation. The white paper reports that the EU is in a strategic position to develop critical improvements in computing power, but it

<sup>20</sup> Roland Berger and France Digitale, “The road to AI: Investment dynamics in the European ecosystem,” (2019), <https://www.rolandberger.com/it/Publications/The-road-to-AI.html>.

<sup>21</sup> Roland Berger, “The road to AI: Investment dynamics in the European ecosystem.”

<sup>22</sup> Perrault, Raymond et al., “Artificial Intelligence Index Report 2019,” (Stanford HAI, 2019), [https://hai.stanford.edu/sites/default/files/ai\\_index\\_2019\\_report.pdf](https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf).

<sup>23</sup> “Sizing the prize: What’s the real value of AI for your business and how can you capitalize?” (PWC, September 2018), <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>.



## PURPOSE: INPUT ON THE EUROPEAN COMMISSION WHITE PAPER “ON ARTIFICIAL INTELLIGENCE – A EUROPEAN APPROACH TO EXCELLENCE AND TRUST”

is crucial that access to computing power – which can be expensive and environmentally costly to acquire – is available for a broad range of institutions and companies within the EU to ensure competitiveness in both industry and expertise. A second prerequisite for the adoption of artificial intelligence is the full digitization of processes of production, with some industries leading the way. One effort that the EU can establish is the development of a policy strategy to foster the digitization of production more comprehensively, focusing on sectors that at this moment are lagging in data, computing power, and digitization.

If the EU hopes to remain competitive in the AI space, it must create policies enabling greater investment in computer chip design and production for AI, as the US holds significant leverage in the marketplace considering its dominance in the global AI chip supply chain. To overcome this gap and become an influential leader in AI technology, it is crucial that the EU improves its investment efforts in AI through public private partnerships. These initiatives must be focused on inclusive innovation and fostering ideas that improve the lives of the many and not the few. For this goal, the EU could institute specialized grant competitions that incentivize and reward inclusive and equitable technological development.

As technical developments soar, algorithms have the potential to replace a great deal of human labor. While these changes suggest many benefits, they will also inevitably condemn many Europeans to face job displacement and unemployment. Not only will workers be displaced from tasks they previously performed, but there will also be a decrease in demand for lower-skilled labor as industries develop due to the increases of automation and AI. “The extent to which these technologies displace workers will depend on the pace of their development and adoption, economic growth, and growth in demand for work. Even as it causes

declines in some occupations, automation will change many more – 60 percent of occupations have at least 30 percent of constituent work activities that could be automated.”<sup>24</sup> Yet, automation may also create millions of new jobs globally, and the equilibrium between creation and displacement will be critically affected by the labor policies the EU will implement.

### 3.2 JOBS AND SECTORS ADOPTING INCREASED AUTOMATION

Increased automation not only threatens Europeans’ access to employment, but will also force European economic and political institutions to come to terms with supporting a larger number of citizens who lack the necessary training and skills to take on new economic opportunities. While some sectors may absorb this shock, one prediction suggests that “globally, up to 375 million workers may need to switch occupational categories.”<sup>25</sup> This was estimated before the COVID-19 crisis, which will place additional pressure on job markets and government finances. These changes will impact different countries asymmetrically and at different times, as some are more reliant on industries soon to be impacted. The EU can adopt policies that will accommodate new AI-based occupations by committing to the development of education and training programs that will best enable humans to work in conjunction with machines, and increase skill sets for AI-enabled jobs. It could also ensure transferable benefits and commit to increasing support for those navigating the future of the labor market such as those EU member states that have already begun testing ideas like universal basic income and adaptive social safety nets.

Employment decisions and workforce management,

<sup>24</sup> McKinsey Global Institute, “Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation,” (McKinsey & Company, December 2017), <https://www.mckinsey.com/-/media/mckinsey/featured%20insights/Future%20of%20Organizations/What%20the%20future%20of%20work%20will%20mean%20for%20jobs%20skills%20and%20wages/MGI-Jobs-Lost-Jobs-Gained-Report-December-6-2017.ashx>.

<sup>25</sup> Ibid.





## PURPOSE: INPUT ON THE EUROPEAN COMMISSION WHITE PAPER “ON ARTIFICIAL INTELLIGENCE – A EUROPEAN APPROACH TO EXCELLENCE AND TRUST”

particularly in the service sector, is an additional area for potential non-physical harms as a result of AI implementation. Some companies in the U.S. are already using AI systems to pre-screen candidates’ online profiles and extract personality information. Others may use AI to discourage some candidates from applying by selectively advertising their job postings on aggregate job sites. These and other forms of online discrimination should be carefully considered by EU regulators as a potential threat to the EU Single Market foundations as well as to fundamental rights. In the service sector, AI is being used for increased monitoring, real-time data analytics, and automatic business adjustments. Some will seek to automate decisions that apply to workforce management, such as promotion and termination. In order to protect the rights of workers on the job, it is necessary that the European Union develop a system of regulation of the application of AI technologies for employment and management decisions. Monitoring of workers should not impede their right to privacy. Regulation may also include a requirement for a “human in the loop” for AI employment and human resource decisions.

The expansion of AI and the inclusion of automation in many other fields is seen as potentially threatening to the current labor market, and specifically, to jobs that are more susceptible to being replaced by such technologies. This concern is certainly valid, as in many contexts AI is developed with the explicit goal to substitute machine work for laborious human tasks. In some of these contexts, the AI applications serve to reduce error and human risk (e.g., advances in health testing or recycling), but, in others, robots may simply replace bodies. “Almost one-fifth of the time spent in US workplaces involves predictable physical activity and is prevalent in such sectors as manufacturing and retail. These sectors have a relatively high potential for automation given the capabilities of current AI technologies. Even within sectors, there is considerable variation. In manufacturing, for example,

occupations that have a large proportion of physical activities in predictable environments such as factory welders have a technical automation potential above 90 percent, whereas for customer service representatives that potential is less than 30 percent.”<sup>26</sup> The same is true for these sectors in Europe. It is therefore important that AI technologies are developed with the intent of innovating and enhancing the current working environment, solving for the creation of new job positions within the same industry in which the technology is adopted, or ways to retrain any displaced workforce.

“The changes in net occupational growth or decline imply that a very large number of people may need to shift occupational categories and learn new skills in the years ahead. The shift could be on a scale not seen since the transition of the labor force out of agriculture in the early 1900s in the United States and Europe, and more recently in China. But unlike those earlier transitions, in which young people left farms and moved to cities for industrial jobs, the challenge, especially in advanced economies, will be to retrain midcareer workers. There are few precedents in which societies have successfully retrained such large numbers of people. Frictions in the labor markets—including cultural norms regarding gender stereotypes in work and geographic mismatches between workers and jobs—could also impede the transition.”<sup>27</sup>

A major concern with new AI technologies is their potential to exacerbate inequality and generate division within political and social systems. It is important to remember that there are many jobs that are impossible to automate and that will be in higher demand in the future, such as jobs that require personal communication, empathy, and creativity. Within the framework of the EU White Paper, the European Commission should include a specific goal to bolster the development of such skills and expand education programs that focus on these skills.

<sup>26</sup> McKinsey Global Institute, “Jobs Lost, Jobs Gained.”

<sup>27</sup> McKinsey Global Institute, “Jobs Lost, Jobs Gained.”



### 3.3 SKILLS REQUIRED FOR AI EXPANSION

As highlighted in the white paper, the adoption of AI technologies requires the extensive use of a large amount of data that requires sophisticated analysis in order to be interpreted. The expansion of AI job opportunities will, therefore, require increased investment in educational programs in universities and education centers, and online, especially for retaining midcareer workers. The training will require newly adapted skills requisite to work in AI, both in the production and operation of AI systems. This would imply a concerted investment from university consortiums in the European Union to expand graduate programs in fields related to AI, not only by replicating successful initiatives but by also expanding current European research grants to attract students who might be otherwise enticed to study outside the EU.

While university initiatives focused on AI might provide a next generation of experts, today’s workforce should also have company-driven training for retooling and AI skills. Care providers, educators, managers and executives, information technology professionals, builders, and more will be impacted. To speed up the adoption of AI technologies by subject matter experts, the EU could develop a series of on-the-job training programs that facilitate the acquisition of AI skills for their workers. The EU faces steep competition for talent from many countries with attractive research and industry leaders, and companies are in competition with larger, international tech giants to attract the most highly skilled workers. This concern is exacerbated in the context of AI, where the qualified workforce is limited even in contexts at the front-end of its development. SMEs will require extensive support from the EU to fully overcome this difficult challenge. We implore the EU to quickly create plans for training for applied AI expertise, interaction with stakeholders, and

transferable management skills related to fields that are rapidly adopting automation throughout the EU.

The European Union should also consider investment in AI projects aimed at extending the working life of the older population and workers with disabilities. In one example from the field of manufacturing, companies such as 99DegreesCustom in the U.S. use highly-developed AI technologies to increase the level of customization available for their products. The customization affords workers within the manufacturing sector with new and more interesting opportunities on the job, requiring creativity over physical strength. AI systems can be a powerful ally to human operators, making certain dangerous or difficult tasks simpler, and ensuring safer working conditions for individuals. Older workers face a higher risk of obsolescence due to the rapid change in the working environment and limited on-the-job educational opportunities offered. The European Union should seek to develop and foster AI technologies that improve the production capacity not just in terms of efficiency, but also in terms of the quality of the production process and end products, incorporating new labor not typical for certain sectors once certain physical demands have been automated.



## 4. Unknown Risks for Widespread AI Adoption in the EU

Non-transparent and unaccountable decision making, as well as potentially biased training data, are outlined as known risks in current systems that employ algorithms for decision making outputs. Gaps in legislation authority and regulation for liability are included in the original white paper as known unknowns. While the paper generally paints artificial intelligence in a positive light, with sweeping beneficial changes for citizens, companies, and the entire European Union, there are three main gaps related to unknown unknowns which require attention for risk analysis and mitigation at the technical level for EU AI policy planning:

### 4.1 UNMARRYING MILITARY AI AND THE BROADER ECONOMIC MARKET

The decision to exclude the development and use of AI for military purposes in the EU AI White Paper limits recognition of the truly dual-use nature of the technology itself and its drivers. It also omits the geopolitical dimensions of the competition for values, interests and standards that are at play at the time of writing this. Defense spending in both research and development for AI, as well as off the shelf commercial products and rapid prototyping, accounts for trillions of dollars spent globally. Many military AI systems are not limited to weapons systems and can be built or adapted for military and non-military use, and many startup companies rely on defense customers to develop, test, and deploy early versions of their products. Many useful tools and techniques have been born from military research and development, including the internet itself. Separating defense and non-defense AI applications presents a missed

opportunity for close cooperation in understanding and regulating what is expected to be the most disruptive dual-use technology in modern history. It also makes it more challenging to ensure the use of AI in military contexts is done in line with democratic oversight and respect for fundamental rights.

Other parts of the world that are advancing plans for the fourth industrial revolution will utilize commercial and military investment and testing interchangeably, buoying strongholds on certain industries, and potentially gaining leads toward faster market entry. Advances in AI are being implemented in non-weapons systems for emergency response, search and rescue, software as a service platform and more. In many cases, diffuse military departments and allies have access to more and better data sets, and/or also shed light on just how much data is yet to be digitized. It would benefit the EU to consider military AI applications in its broader plans, to prevent and avoid fragmentation in the internal market.

The EU has championed much of tech policy to protect the fundamental rights of humans, their data and privacy. However, a lack of prioritization of sector-specific advancements for AI could create a ‘race to the bottom’ scenario in an effort to deploy AI quickly at cost and scale. With \$1.5 billion invested it will be difficult to lead in all areas; “industry, health, transport, finance, agrifood value chains, energy/environment, forestry, earth observation and space.”<sup>28</sup> There is a risk that improvements in explainability and bias for AI systems will take a back seat to market drivers. Decentralized development and testing across member states can be an effective strategy, but it is essential to scope the

<sup>28</sup> European Commission, “EU AI White Paper.”



## PURPOSE: INPUT ON THE EUROPEAN COMMISSION WHITE PAPER “ON ARTIFICIAL INTELLIGENCE – A EUROPEAN APPROACH TO EXCELLENCE AND TRUST”

most important use cases to properly vet the interoperability and regulatory challenges. The EU and its member states cannot effectively prioritize and mitigate risk without prioritizing sectors which are most important for AI over the next 5-10 years, and pairing risk analysis and compliance mechanisms to those priorities, including development of the conformity assessments, providing baselines for future analysis.

Although the white paper mentions the legislative scope of the EU extending to “all relevant economic operators providing AI-enabled products or services in the EU,”<sup>29</sup> it does not address regulating EU exported technology and the possible unintended consequences of systems developed in the EU and used in unforeseen ways outside the EU, both military and non, (e.g. facial recognition purchased from a democratic state deployed for surveillance in an autocratic state). The European Union, home to esteemed international dual-use and export control regimes, has championed success in dynamic trade policy and regulation. The EU now has a unique opportunity to steer the burgeoning field of tech policy compliance, emerging as a parallel to international trade policy, by incorporating global companies with global stakeholders to advance both market economies and regulation authorities on the use of AI.

### 4.2 PREPARING SOCIETY FOR PITFALLS OF AUTOMATION BIAS

In considering ways to improve trust and accountability of artificially intelligent systems, the paper does not address automation bias – the potential for *overtrust* in AI, where human operators expect a certain outcome based on assumptions, and over time, eliminate mechanisms for

quality control. Ongoing research looks at the “tendency of humans to defer to technology when presented with conflicting information,” and the phenomenon’s potential impact to physical security.<sup>30</sup> Without proper training and indoctrination, over reliance on machine outputs can lead to mistakes, and introduces a blind spot in risk management across whole enterprises.

Enterprise adoption of AI systems cannot be naïve to automation bias. “Building an ecosystem of trust is a policy objective in itself and should give citizens the confidence to take up AI applications and gives companies and public organizations the legal certainty to innovate using AI.”<sup>31</sup> The AI value chain must take into account new user interface dynamics when considering the factors surrounding human interaction with AI. It is not enough to test the technological innovations without users and human patterns of behavior. Page 12 of the white paper outlines examples where material and immaterial harm can be exacerbated by automation bias, leading to potentially worse biased outcomes and discrimination in sectors such as law enforcement and the judiciary. The automation bias problem goes beyond the “black box problem” in engineering, and may affect legal thresholds and liability, since current legislation only addresses safety risks at the time a system enters the market, and remains immature on “transparency, traceability and human oversight.”<sup>32</sup> It will be difficult to establish accountability over the lifecycle of a system. A law or regulation must be vague enough to cover different systems deployed across sectors and industries, specific enough to govern lifecycles of systems, updates, and adaptations, despite any explainability gaps.

In promoting the uptake of AI across the private and public sector, not enough emphasis has been placed on the cultural transitions necessary to safely achieve its ambitious goals

<sup>29</sup> European Commission, “EU AI White Paper.”

<sup>30</sup> Alen Wagner, Jason Borenstein, and Ayanna Howard, “Overtrust in the Robotic Age,” *Communications of the ACM*, Vol. 61, No.9, (September 2018), 10.1145/3241365.

<sup>31</sup> European Commission, “EU AI White Paper.”

<sup>32</sup> Ibid



to introduce AI into every facet of human life. The paper mentions attempts to hack or manipulate data or algorithms deployed for intended outcomes, but does not address hacking and manipulation of human behavior and interaction with deployed systems. The Revised Coordinated Plan for AI should expand on requirement type C – ‘information provision’ – to include training on automation bias, dual-use and malicious applications of AI. Many comparisons can be drawn from the development and patterns of behavior catalyzed by the proliferation of the internet, from e-commerce’s explosion as the result of online Beanie Baby sales, or the nefarious use cases of child exploitation and pornography online. As stated, requirement C could be implemented as a type of surgeon general warning, but should be enhanced to a defined body establishing frameworks and best practices to coordinate with the public and end users of systems. EU governance bodies cannot ensure that a deployed system won’t deduce a more efficient way to achieve its goals and objectives in a way uncaptured in its description or confidence intervals, therefore, a governing body must be entrusted to keep pace with developments.

“The plan will also increase awareness of AI at all levels of education in order to prepare citizens for informed decisions that will be increasingly affected by AI.”<sup>33</sup> The Revised Coordinated Plan for AI should consider expanding on this leadership to develop widespread and publicly available training on basic nomenclature, development plans, assumptions, and risks for a general audience. A “curriculum” for developers of AI turned into training materials is not the same product for promoting tech literacy for the whole of Europe. The plan should consider dialogues which in turn bring technologists out of research labs and into communities. Consider for example an AI developer whose technology for autonomous vehicles struggles to identify small dark objects visiting a predominantly African-heritage neighborhood, to grapple with the reality of that system

deployed in that environment which might not be able to identify children.

### 4.3 TESTING AND OPERATING IN MULTI-SYSTEMS ENVIRONMENTS

Machine learning systems which are programmed with different rules based on different values, dependent on different inputs, risk calculations and tradeoffs, have not been deployed at scale in the same environment. Discussion of testing for AI systems often cite two phases, testing in environments where the system has been trained to operate (control), and testing in environments where the system has been trained to learn based on controls. What is often overlooked is environments which deploy multiple different artificial intelligent systems, for example, testing various autonomous vehicle systems operating simultaneously on the same roads. This “systems on systems” black hole extends to military applications, health care, finance, language, predictive analysis, etc. and highlights the unique “many hands” interoperability problem.

Systems with the same objectives could have different outputs, different biases, or one or both could lack nuanced data. What happens when a linear system has to communicate with a vector-type system, or a vector-type system needs to be interoperable with a decision tree system? Can one system account for concept drift, a statistical scenario where “our interpretation of the data changes with time even while the general distribution of the data does not,”<sup>34</sup> in another? Supervised systems interacting with unsupervised systems? Integration of AI systems will not occur in a vacuum. As one researcher illustrates a portion of this problem, “rather like having the Lakers play the Patriots in the World Series, when both the concept/game and the data/

<sup>33</sup> European Commission, “EU AI White Paper.”

<sup>34</sup> Ashok Chilakapati, “Concept Drift and Model Decay in Machine Learning,” *Towards Data Science*, April 25, 2019, <https://towardsdatascience.com/concept-drift-and-model-decay-in-machine-learning-a98a809ea8d4>.



## PURPOSE: INPUT ON THE EUROPEAN COMMISSION WHITE PAPER “ON ARTIFICIAL INTELLIGENCE – A EUROPEAN APPROACH TO EXCELLENCE AND TRUST”

players change there would be a lot of head scratching in the stands.”<sup>35</sup> While systems on systems is clearly a technical and meta problem, the Revised Coordinated Plan for AI should provide a step by step approach to testing systems which will inevitably operate in the same environment. This area of study would be a boon for the member states’ Digital Innovation Hubs and could appoint sector-specific research to each and entice talent to procure, train, and test contrasting models in a sector-specific environment.

There is a need for tiered testing phases to include testing systems on systems given that mistakes and biases can result from AI learning in operation. AI systems may also learn from or contrast with other AI systems. This phase of testing is essential, “where the outcome could not have been prevented or anticipated at the design phase, the risks will not stem from a flaw in the original design of the system but rather from the practical impacts of the correlation or patterns that the system identifies in a large data set.”<sup>36</sup> This requirement could be developed into its own conformity assessment cycle prior to deployment, meeting requirements for addressees in a more complex testing environment and raising the bar for both security by design and risk prevention.

It is necessary to revalidate the European Union’s ability to interoperate digitally in order to execute combined military tests and operations. Asynchronous capabilities currently exist among member states; therefore, a standard for the use of AI in military operations must be established to ensure that new developments do not inhibit interoperability. In the case of AI development for military weapons systems, a clear delineation should be drawn between offensive and defensive systems. The EU should conduct a study to determine the areas where increased automation would benefit or weaken infrastructure. The implementation of AI in defensive capabilities – known as Intrusion Detection Monitoring (IDM) systems, should be standardized. The implementation of this

capability has the potential to increase defense posture, limit miscalculation, and deter intrusions. Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) capabilities are susceptible to cyber enabled effects. This should not deter from the use of AI in C4ISR for offense, but rather highlights the need to validate C4ISR targeting data prior to its use in operational decisions. This same susceptibility implies that the additional risk introduced by AI in offensive platforms’ where deployment and execution may exceed moderate risk tolerance.

Where offensive cyber-enabled capabilities can be enabled through AI, human in the loop decision responsibility must be enshrined. In order to maintain control over military AI tools it is recommended that the EU nest this capability and outline systems intent with an international framework, akin to the Tallinn Manual. This will provide a clear declaration of what is and is not acceptable for deployment in the military context, enhancing authority over the development of AI defense trajectories. AI-enabled offensive weapons, if developed, should abide by existing international laws and norms. It is further recommended that AI-enabled capabilities remain egalitarian and distributed within the EU. This will ensure that these capabilities do not hinder military operations in the absence of an EU AI military implementation strategy. Early risk assessment measures here are also a must.

<sup>35</sup> Chilakapati, “Concept Drift and Model Decay in Machine Learning”.

<sup>36</sup> European Commission, “EU AI White Paper.”



## 5. Conclusion

As research and development produce more capable domain-specific AI systems, it is likely that the resulting economic effects will have significant impact on populations, markets, and politico-economic relations between states. Advances in robotics coupled with an increasing consumer preference for quick delivery of customized products might reduce labor costs and shift manufacturing toward localized economies. This may pose challenges for member states highly dependent on manufacturing exports, increasing states’ dependence on natural resources and deepening inequality within and between states.

These concerns are not limited to physical goods. As machine learning becomes more sophisticated, it is likely to bring disproportionate financial gains to internet companies that already possess large amounts of data. These firms will then have the greatest capacity to hire, cultivate, and retain talent, potentially creating a small group of firms which own a lion’s share of gains from AI. As a result, these firms could become increasingly important to their relative states, both because of national competition for “algorithmic supremacy,” but also for the tax revenue these companies generate. States without powerful AI firms could find themselves increasingly dependent on other states, which may find themselves exceedingly dependent on powerful companies.

Aside from concerns about power, matters of equity and fairness have become abundantly clear when considering futures where technological prowess is concentrated in just a few centers. Given the diversity within and between EU member states, where to distribute headquarters for AI development and regulation remains an open question. Could a small number of member states be relied upon to serve as responsible stewards for the specific needs and values we have identified? States and their constituencies may have different preferences regarding tradeoffs between fairness and

accuracy, or privacy and productivity, which may not be easily addressed at the EU level.

By no small effort, we invite the European Commission to consider the comments and recommendations made throughout as those of conscious global citizens. We want to see the EU promote the transparent and equitable distribution of AI research, development and deployment. We encourage initiatives to increase public awareness, training, and literacy in response to advancements in AI, and suggest the creation of new occupations in the data-driven future. We also hope for increased understanding of and export control over military machinery and proliferation pathways for dual-use technology. These recommendations can be coordinated and operationalized throughout the EU, made up of distinguished interdisciplinary experts, to tackle the implementation of dynamic policies as they relate to the development and trade of AI hardware and software, cooperation, and the capacity for change. We submit these recommendations for your consideration, and look forward to the European Commission’s comments.