Palantir Technologies' Response to the EC's Consultation on the White Paper on AI

1. INTRODUCTION

Palantir Technologies is a software company with a global presence across multiple countries, many of which are in the European Union ("EU"). We build data platforms that enable public and private institutions to integrate, analyze, and collaborate on their data in a privacy-protective way. Our vision is a future in which public institutions, commercial enterprises, and non-profit organizations are fully equipped to more effectively and responsibly use their data to carry out their mandated goals, to deliver value to their customers, and to provide critical services to those most in need. As we build and implement technology, we believe that protecting fundamental rights and freedoms is essential to that mission.

We welcome the European Commission's ("EC") commitment to a regulatory and investment-oriented approach with the twin objective of promoting the uptake of artificial intelligence ("AI") and addressing the risks associated with certain uses of this new technology. We want to contribute to these efforts by sharing some of the lessons about data integration and analytics we have learnt in our 15 years history. Specifically, we want to describe some of the principles that we have developed internally as our guidance for responsible and value-enhancing use of information assets, including and especially as they relate to advanced data science and analytics techniques.

Our experience of working with a broad range of private and public organizations has allowed us to gather unique pragmatically oriented insights into the challenges of responsible data integration and analytics as the foundations of AI systems, but also of AI itself, as it is becoming an increasingly important extension of our offering. We believe that some of these lessons can be a helpful resource for the European Commission ("EC") as it works towards the policies of the future European AI ecosystem of excellence and trust.

Given Palantir's strategic and operational focus on responsible data integration and analytics, our response to the Commission's White Paper On Artificial Intelligence – A European approach of excellence and trust published on February 19, 2020 ("White Paper") is limited to Section 2 of the questionnaire.

2. RESPONSES TO SECTION 2 OF THE QUESTIONNAIRE: AN ECOSYSTEM OF TRUST

2.1. CONCERNS AROUND AI

The section below summarizes our thinking about the concerns around AI and proposes several technical, administrative and policy strategies for the EC to consider. The

concerns we have identified (in addition to those already listed in the White Paper) are: proliferation of stakeholders, the difficulties with controlling AI, the aggregate impact of AI, and errant framings of AI as a panacea to socially and culturally situated problems of irreducible complexity.

2.1.1. Proliferation of stakeholders

AI systems' life cycle typically involves multiple stakeholders who all bear responsibilities for safe handling of AI:

- developers (those who research, design, and/or develop AI systems),
- data scientists (who build, train, and refine the algorithms powering AI),
- deployers (public or private organisations that use AI systems within their business processes and/or offer products and services to others),
- end-users (those engaging with the AI system, directly or indirectly), and
- broader society (those directly or indirectly affected by AI systems).

The High-Level Expert Group on AI set up by the EC ("HLEG"), a source that the White Paper refers to at several points, has proposed a set of 14 requirements for trustworthy AI such as transparency, robust security, privacy, human agency, and accountability that should be applied throughout the AI systems' lifecycle.[1] According to the HLEG, developers (and, we should add, data scientists, although this category is not explicitly mentioned in the HLEG's guidelines) should implement and apply the specified requirements to design and development processes.[2] Deployers should ensure that the systems they use and the products and services they offer meet the requirements. End-users and the broader society should be informed about these requirements and have the ability to request that they are upheld.

However, there might be a gap between the vision outlined in the HLEG's document and business reality (which we, as practitioners operating in this space, have had direct experience navigating and therefore feel uniquely qualified to comment on). Not all the stakeholders follow the same standards of care, possess the same level of technical skills or receive the same incentives to remain accountable. More importantly, stakeholders may differ in their approach to ethics and their value systems. This proliferation and diversity of stakeholders' practices in the development and use of AI unavoidably increases the risk of potential harm caused by AI systems.

Palantir helps customers build an infrastructure that supports modeling, evaluation, and deployment of AI applications. This mode of engagement often entails working closely with our partners in various stages of the deployment of AI that is specific to their problem sets. In this position we have recognized that all phases of AI development and deployment can be subject to inherent limitations (e.g., personal bias, limited understanding) and extrinsic failings (e.g., poor engineering practices). As already mentioned, these limitations and failings may be accentuated due to proliferation of

[1] The High-Level Expert Group on AI, 'The Ethics Guidelines for Trustworthy Artificial Intelligence (AI)' (2019) <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top> accessed on 25 May 2020.
[2] *Ibid*., p. 14.

stakeholders.

When our customers employ AI/ML tools in our platforms, Palantir as technology provider recognizes the need to adopt a more holistic approach through enabling our sofware's use in a way that provides a connecting tissue between the stakeholders to ensure not only the viability and efficacy of the technologies themselves, but also the fidelity and fairness of the data and AI use. For example, we often share data governance principles with the deployers and designers of AI systems, educate end-users and address concerns together with civil society representatives. This unique position as a data integration platform provider has enabled us to form valuable relationships with multiple stakeholders including designers, users and deployers, and has been critical for ensuring trust and protecting values in the use of our software. Based on our positive experiences, we encourage the European Commission to consider policies that would support and incentivize enhanced cooperation and partnerships across the AI lifecycle.

2.1.2. The difficulties with controlling AI

The technical and administrative set-up of the environments in which AI is deployed is key to understanding the risk. For example, is it possible to see and correct information the AI is using about individuals in a straightforward manner? Is there a right to have the system reviewed by a human?

Imposing control and ensuring oversight are two difficulties that are common to many AI systems. According to the HLEG, human agency and oversight in various forms should represent a critical feature in the system of trustworthy AI.[3] However, human supervision of AI – in any form – must be appropriately planned for and embedded at the outset by designing AI systems to include technical measures and processes that enable proportionate measures of human intervention and control.

Palantir has over 15 years of experience of dealing with governance around information resources in many sensitive environments (i.e., information assets including analyses, visualizations, AI/ML models, data transformations, data pipelines, etc.) and has acquired deep expertise in enabling the institutions using our platforms to effectively enforce accountability and human supervision through a number of discrete measures and techniques. System usage auditing has been at the forefront of these efforts and has revolved around two types of techniques:

i. in-process system records that catch unintentional or malicious privilege escalation, and
ii. post-incident analysis of audit logs to interrogate and document events contributing to specific outcomes at later stages.[4]

---

[3] *Ibid*., p. 16.
[4] Mark Elliot, Robert Fink 'Technical Challenges in Data Sharing' *Medium* (10 April 2020) <https://medium.com/palantir/technical-challenges-in-data-sharing-3ec353f6c4da>; see also Courtney Bowman, 'Data Protection in Palantir Foundry' *Medium* (22 April 2020) <https://medium.com/palantir/data-protection-in-palantir-foundry-5ab9f346195> accessed on 25 May 2020.

Having realized how helpful the audit practice has been for supporting our customers in carrying out responsible and comprehensive oversight and accountability measures, we encourage the European Commission to consider information system (including AI/ML tools) usage auditing as an inherent part of data users' accountability in relation to processing of data inside and outside of the GDPR scope.

2.1.3. The cumulative impact of AI

To accurately assess the impact of AI, the evaluation should not necessarily focus on an individual product, but it should focus more on considering the impact of the cumulative or scaled deployment of the product to society. For instance, a single Ring doorbell camera with facial recognition is less concerning than a neighborhood networked with such cameras that could allow someone to track individuals as they traverse the area.[5]

2.1.4. AI is not a panacea

While it may sound disappointing to technology optimists, it is necessary to accept that, at times, AI is not the appropriate or optimal response to the challenges in the business and/or public sectors. We believe the White Paper and its readers would benefit from more explicitly addressing threshold limitations of AI and further articulating characteristics and considerations of areas of speculative application where AI systems are, in fact, simply ill-suited solutions and should be avoided. (Readily deployable) AI systems tend to hit their limits when they are applied on complex social behavior and when adequate mitigation measures may not be able to mitigate risks of addressing such nuance and complexity to an acceptable degree. For example, applying Natural Language Processing on social media messages to analyze the richness of sentiment in complex speech environments appears scientifically uncertain as well as morally and legally questionable.

2.2. THE NEED FOR NEW RULES FOR AI SYSTEMS

While the existing EU framework around data (particularly the GDPR and human rights mechanisms) may have certain gaps, its principle-based nature makes it fairly comprehensive. Thus, more than potential gaps, the concern is that the framework is not sufficiently translatable to be fully applied in practice. To address the issue, the HLEG has proposed to leverage internal procedures and policies aimed at securing compliance with data protection laws. These existing legal procedures do indeed encourage better operationalization and contextualization that in turn help facilitate ethical data handling. Over the past years, Palantir has invested considerable intellectual and other resources into breaking down the regulatory principles and rules into practical operational guidance for dealing with the challenges of AI systems within and beyond legal compliance. These are some examples taken from our internal policies/processes that we hope will be useful for the EC to consider:

---

[5] The Ring Video Doorbell, a Ring Inc.'s flagship product, is a smart doorbell that contains a high-definition camera, a motion sensor, and a microphone and speaker for two-way audio communication. 'Ring Inc.' *Wikipedia* (2020) <https://en.wikipedia.org/?title=Rush> accessed on 25 May 2020.

- building up a set of data governance principles that protect data misuses on the most fundamental level such as mandatory role- and purpose-based access controls;[6]
- taking a strong proactive stance towards specific applications of data and AI techniques, such as facial recognition and other biometrics;[7] and
- representing data in a more human-intuitive fashion that increases transparency and accountability around decision-making processes.[8]

2.3. MOST CONCERNING ("HIGH-RISK") USES OF AI

The Commission is of the opinion that a given AI application should generally be considered high-risk in light of what is at stake, considering both the sector and the intended use. As a general comment to the question above, Palantir agrees that context matters enormously and is essential to determining whether specific types of information should be used in AI, and in some cases whether AI should be deployed for a given application in the first place. For example, it is one thing to seek to analyze whether sensitive categories such as race or gender linked to genetic attributes correlate with certain disease prevalences. It is quite another thing to seek an association between those same immutable personal attributes and one's likelihood to default on a loan.

High-risk sectors and/or uses require particularly attentive handling of AI, as it is more likely that those AI models will include sensitive personal attributes that have real impact on people's lives. Therefore, if such sensitive categories are implicated in AI programs either by necessity or choice, their inclusion should be identified and, where appropriate, contextually justified in clear, accountable terms. Unfortunately, sensitive features are not always easily identified. For example, names and addresses would not typically be considered sensitive, but a combination of a foreign surname and residency in an economically disadvantage neighborhood may serve as a proxy for ethnic and financial status. Organisations dealing with high-risk sectors and/or uses can mitigate the risk by implementing the following solutions:

1. a meticulous evaluation of how the model depends on potentially discriminatory features and their proxies, knowing that discrimination is not a single concept that is easily agreed upon and that fairness metrics may conflict with each other, and
2. a framework for the periodic review of semi-static adjustments to the feature selection and weights in the model and the underlying data.

The Commission could consider these solutions as some sort of soft policy measures to incentivize a more structured and thorough approach to addressing this class of AI risks.

---

[6] *Supra* 4.
[7] Bob Gourley, 'Insights for your enterprise approach to AI and ML ethics: Context from Courtney Bowman and Anthony Bak of Palantir' *CTO Vision* (26 March 2019) <https://ctovision.com/insights-for-your-enterprise-approach-to-ai-and-ml-ethics-context-from-courtney-bowman-and-anthony-bak-of-palantir/> accessed 25 May 2020.
[8] Paula Kift, 'Augmentation as artifice: a Palantir look at AI' *about:intel* <https://aboutintel.eu/palantir-augmented-intelligence/> accessed 25 May 2020.

2.4. SPECIFIC MANDATORY REQUIREMENTS OF A POSSIBLE FUTURE REGULATORY FRAMEWORK

From the perspective of a company that provides data integration platforms, we have a deep appreciation for how an understanding of the provenance, pedigree, lineage, and quality of training data, as well as record-keeping around derivative model development and refinement are two of the most critical but often overlooked requirements of a possible future regulatory framework for AI, because any weakness in a data asset may be reflected in the AI system built on it. The White Paper does acknowledge the importance of the training data quality, but we believe this aspect would benefit from additional attention and analysis. Especially for high risk situations, a more granular approach to data assessessments and tracking could be helpful, breaking the concept down into at least a few components:

- Accuracy of the data (For example, is the data labeled correctly? Are the features correct?);
- Data completeness (For example, does the data cover all the examples, for instance are all races, ages, genders etc., represented for all possible labels?); and
- Provenance and lineage of data assets (For example, where does the data come from? How has it been modified or changed over time?).

Although data serves as a foundational component of all AI systems, it is important to keep in mind that data in and of itself does not rise to the level of a usable AI asset. In order to be useful in developing new or augmenting existing AI systems, data requires tremendous human labor to catalog, label, vet, examine, etc. In spite of how much it is needed, this work often gets overlooked and treated as unnecessary and/or too expensive.

The White Paper proposes amendments to the liability directive that would lead to better safety of AI-driven products through regulating data quality. In this regard, Palantir wants to draw attention to its set of data governance practices which highlights some of the key industry concepts to ensure data trustworthiness.[9] For example, organisations should be encouraged (by regulatory measures if there are not sufficient economic incentives) to invest in (i) data management practices and supporting infrastructure and (ii) security around data. The former centers around an ability to evaluate data quality, to curate it with critical contextualizing details, and to ensure its recency. The latter suggests that, to the greatest extent possible, technology should safely enable data sharing only to the extent needed and, critically, only for the time period needed. Data governance not only contributes to safe, secure, and restricted application of AI technologies, but has benefits beyond that, for example in relation to privacy and transparency concerns.

2.5. FURTHER EU-LEVEL GUIDELINES FOR HIGH-RISK TECHNOLOGIES

At the outset, we note that biometric identification systems such as facial recognition ("FR") can be applied in a variety of ways, ranging from less intrusive technologies (such as local, one-to-one face authentication) to remote face recognition technologies that

[9] *Supra* 2.

could effectively create a ubiquitous surveillance system.[10] Thus, just like any other technology, AI systems may potentially be very dangerous if no robust privacy and other safeguards are put in place. We applaud the EC for articulating the concern in the White Paper according to which AI may be restricted for remote biometric identification purposes where such use is duly justified, proportionate, and subject to adequate safeguards.

Using the example of FR, we want to further elaborate on the idea of proportionality and adequate safeguards that should be built into the AI systems.

As a starting point, whenever FR is being considered, the adopters need to clearly define what is the problem the technology should be trying to solve. This mental exercise will enable them to investigate if less intrusive solutions could be put in place. For example, in many cases the less intrusive facial authentication will be as sufficient as remote facial recognition solutions.

Particularly for those more sensitive applications, we propose the use of 'an obfuscation by design' technique (ObD). ObD means that in cases when the facial recognition technology detects faces that are (a) known but do not trigger an alert or (b) are unidentified, the end users are only exposed to blurred imagery. The success of this technique goes hand in hand with the implementation of some key GDPR data protection principles such as storage and purpose limitation, data minimization, accuracy, and accountability.

2.6. FURTHER SUGGESTIONS FOR THE ASSESSMENT OF COMPLIANCE

The White Paper suggests a combination of *ex ante* (conformity assessment) and *ex post* compliance assessments. The former would be executed by AI developers and deployers, whereas the *ex post* assessment would require an engagement of a competent authority.

What the White Paper misses is the need to engage external oversight authorities such as relevant regulatory authorities also at early stages to provide topical insights and thereby mitigate risks of adverse regulatory actions. For example, in 2019 the UK Information Commissioner introduced the Sandbox service to support organisations who are developing products and services that use personal data in innovative and safe ways. Context-permitting, institutions deploying AI programs should also consider engaging with the public and other community stakeholders, such as advocacy and civil society groups, as early as possible. This early community engagement can help to demystify the intended applications, avoid unwarranted suspicions, provide an opportunity for stakeholders to voice their concerns, and enable pathways for exploring whether and how those concerns may be addressed.

[10] Wojciech Wiewiórowski, 'AI and Facial Recognition: Challenges and Opportunities' (21 February 2020) <https://edps.europa.eu/press-publications/press-news/blog/ai-and-facial-recognition-challenges-and-opportunities_en> accessed on 25 May 2020.

3. CONCLUSION

AI, like other powerful information technologies, has great potential for the development of our economies and societies, but also poses some serious risks. We firmly believe that AI advances do not need to come at the expense of privacy and other fundamental rights. Our responses to the EC's White Paper propose several technical, administrative, and policy strategies for a balanced use of AI systems. We assert that these strategies are feasible, and we have demonstrated their feasibility in variations of software that our clients use on the grounds. We encourage the Commission to consider our recommendations as the informed perspective of technologists who wish to see future technology developments directed towards socially responsible and defensible uses.