# Submission to the House of Lords AI in Weapons Systems Committee: Inquiry on AI in Weapons Systems

**Palantir Technologies UK, Ltd.**

**April 2023**

## Introduction

1. We appreciate the opportunity to provide evidence to the House of Lords' AI in Weapons Systems Committee. In this submission, we draw upon our experience providing the kinds of software capabilities that underpin military applications of artificial intelligence ("AI") and autonomy. After introducing Palantir, we describe the principles that guide our product engineering and deployments in this field, and then answer the Committee's specific questions.

2. A key theme of our submission is that AI and autonomy capabilities need to be understood in their wider operational and normative contexts. This is especially important in the case of military applications of AI, and even more so for autonomous military capabilities. For this reason, we would encourage the Committee to consider engaging with live demonstrations of the kinds of AI capabilities its inquiry is focused on. As regards our own capabilities, we would be very happy to arrange this.

## About Palantir

3. Palantir Technologies ("Palantir") is a software company that builds data integration and analytics platforms for public, private, and non-profit organisations. Though US-headquartered, London is home to our largest office and primary product development centre, with over 900 employees.

4. The company was founded in 2003, with an initial focus on enabling defence and national security institutions to manage and utilise their data assets more effectively and responsibly. Beyond the defence and national security sector (still an important part of our business), our software platforms are now deployed across every field of industry and government.

5. British, American, and allied armed forces use our software platforms to integrate disparate data – of different formats and source systems, and often highly-sensitive in nature – and then apply this data across various strategic and tactical military functions, while enabling adherence to security, legal, and ethical requirements, as well as applicable norms of conduct. Our platforms support use cases ranging from the support and maintenance of defence platforms (e.g., warships) and personnel management, to intelligence, surveillance, target acquisition and reconnaissance ("ISTAR") workflows. AI capabilities, enabling varying degrees of augmentation and autonomy, feature across these use-cases, leveraging the underlying data infrastructure that our software provides.

## Guiding principle: AI technologies need to be understood in their operational and systems context

6. As a software company, we believe that it is critical to develop software and systems that are informed by operational realities and reflect the constraints and limitations — technological, procedural, and normative — that our partners face in the field. This grounded approach provides a safeguard against much of the confusion and hype that we observe as frequently undermining the fidelity and utility of policy discussions related to advanced technologies. The perspectives we present here are based on that practical experience.

7. Core to that experience is a recurring affirmation of our foundational commitment as a company: that technology must fulfil both the mission needs and the ethical imperatives of its users. Whether this threshold is met cannot be assessed by viewing any single technology, particularly an autonomous weapons system or AI capability, in isolation. Rather we must consider the integrated hardware/software system, where autonomy in one instance might feed into a human decision in the next. For example, in a targeting workflow, a user might task a sensor to automatically gather additional information after a computer vision tool has first identified a potential target in satellite data, but then pass this unified view, including the underlying image and other information, to a human for subsequent decision making. Further, these kinds of integrated hardware-software-human operator systems are themselves situated within the context of broader operational processes and specific institutional requirements (law, doctrine, guidance, best practices, etc.) which will ultimately determine both the integrity of the system and the accountability of the human decision makers within it.

8. British and allied military personnel need effective tools for making informed decisions, deterring near-peer adversary threats, and prevailing in modern conflict. Equally, these tools should uphold the moral traditions and imperatives that make our societies worth defending – they should, for example, help personnel minimise civilian harm and uphold obligations under International Humanitarian Law ("IHL"), and ensure that life and death decisions are being made by appropriate, accountable actors.

9. These considerations are playing out in adversarial contexts in which there is likely emerging — if there is not already — a race towards the development and adoption of advanced technological capabilities. However, it is core to who we are (in contradistinction to our adversaries) that the rule of law (i.e., IHL) is at all times binding and that we steadfastly adhere to long-held values and traditions of conduct (e.g., the Just War tradition). These commitments fundamentally preclude the posture of writing a blank check for technological advancement at all costs. Rather, they reaffirm how important it is that we are successful on our own principled terms.

**Question 1: What do you understand by the term autonomous weapons system (AWS)? Should the UK adopt an operative definition of AWS?**

10. In our view, the role of autonomy – that is, the minimisation or removal of direct human enactment of or supervision over certain tasks or actions – is primarily intended to accelerate or optimise decision-making or operational tasking in relation to existing military capabilities. It follows that we view autonomous technologies, including AWS, more as a methodological grouping concept than as a distinct weapons classification analogous to nuclear, biological or chemical weapons.

11. And yet, because AWS bring into question the role of the human decision-maker in selecting and engaging targets, the category does motivate special and separate consideration from other forms of weapons systems. In our view, that consideration (by governments, civil society, and technology firms) would be aided by a common definition of what constitutes AWS.

12. Specifically, we advocate for a functional definition, such as that provided by US Department of Defense Directive 3000.09: "A weapon system that, once activated, can select and engage targets without further intervention by an operator." This definition may be imperfect (for example, it seems to focus on full autonomy at the expense of considering the spectrum of semi-autonomous weapons that may trigger many similar considerations), but is a directionally useful starting point.

13. By focussing on the most operationally salient functions of AWS, and not on specific technology components, this definition will not readily succumb to technological obsolescence. It is also inclusive of operator-supervised and unsupervised AWS, AWS focused on platform or installation defence, as well as of AWS focused on selecting and engaging targets in other contexts. Further, it leaves open the prospect of context-appropriate human interventions after a system has been activated, for example, overriding a targeting system after it has been initiated based on the

sudden availability of new information, such as the heretofore unknown presence of civilians within targeted infrastructure.

14. A further advantage is that it avoids anthropomorphising or imparting sophisticated human decision-making characteristics to the AWS, instead focussing on the specific tasks that are accomplished without or with minimal human intervention after activation. Attempts to impart "reasoning", "thinking", "judging", "understanding", or other cognitive characteristics to AI or AWS are not justified by the current or foreseeable future state of the technology, and can get caught up in academic back-and-forth trying to define "reasoning" or "intelligence" itself. Instead, we advise treating AI in AWS (and in general) not as intelligence as such, but as classes of tools or capabilities that operate by executing computational functions. That is, AI should not be considered as generating any form of agent-driven intelligence, but rather as computational capabilities that programmatically encode rules, logic, mathematical calculations, or statistical predictions based on training data to carry out designated functions. Even the most advanced areas of AI development – leveraging, for example, neural networks or Generative AI – reduce to computational functions over digital data with no conscious sense of anything like an intentional focus or meaning-bound awareness of data related to things in the world. These types of machine operations may be helpful for augmenting or carrying out specific functions, but they fundamentally remain a far cry from instantiating the complexities of socially situated, physically embodied, culturally inflected human understanding and intelligence that is required for carrying out the most consequential dispositions and decisions that impact human lives and the world at large.

15. A functional definition, by contrast, centres our attention on the primary reason that AWS need special consideration: the potential removal of a human decision maker from the act of selecting and engaging specific targets. Consequently, this may motivate the need for further guidance around questions of predictability, reliability, and explainability of any decisions that AWS under this definition might make, and whether such decisions are appropriate regardless of their predictability, reliability, and explainability.

16. However, we think that any definition of AWS that focusses attention exclusively on full autonomy is problematic. Such a definition risks deprioritising critical considerations that should be carefully addressed when building elements of autonomy into workflows that enable the activation of weapons systems – whether or not those weapons systems fall into a semi- or fully-autonomous category. The introduction of autonomy into military decision making generally — and military targeting specifically — raises many of the same critical ethical and operational questions, whether or not the autonomous function is part of a physically or logically individuated component of a broader system. And while the advent of fully Autonomous Weapons Systems that may be intended to target individuals is a concern for future warfighting, elements of autonomy are with us already and are actively being used to help inform or drive decision making on the front lines, just as active protection systems already have a role in defending ships and installations.

17. It is certainly worthwhile for the United Kingdom to adopt a definition of AWS around which it can continue to develop thoughtful policies and rails. Not only will this be valuable for a whole of Government effort within the United Kingdom, it can also help to elevate the United Kingdom's voice in the international arena on these critical issues. Common international understanding will require at least some consistency of language, and the United Kingdom has a vital role to play in pushing these conversations forward and avoiding a "race to the bottom" when it comes to potential development and deployment of these technologies.

**Question 2: What are the possible challenges, risks, benefits and ethical concerns of AWS? How would AWS change the makeup of defence forces and the nature of combat?**

18. Autonomy, short of fully autonomous weapons systems, is already changing the makeup of defence forces and the nature of combat. In particular, autonomy calls attention to the fact that perhaps the most valuable resource in modern warfighting is time. When deployed effectively, as

part of a human-in-the-loop decision-making process, AI and autonomy increase the time available for critical human-driven decisions, while at the same time reducing the overall time-to-decision and improving the accuracy and reliability of decision-making.

19. For instance, rather than having human analysts pour over satellite images to determine enemy force disposition, AI tools can help automate the identification and prioritisation of relevant potential matches for further human review. Even at this early stage in the workflow (whether for targeting, intelligence gathering, or other), thoughtful management of AI development and deployment is critical - ensuring that the appropriate AI models are deployed following thorough testing and evaluation protocols, that they are consistently maintained, and that their strengths and limitations are understood by the human user.

20. More than that, sophisticated automated techniques can fuse that information (i.e., the satellite imagery) with other sources of intelligence, to almost-instantly provide a holistic view of objects of interests that could otherwise take a human analyst hours to prepare. Again, there are critical ethical considerations in the specific implementation. For example, user interfaces and the wider user experience of a software platform must be designed such that representations of the objects of interest (which may have been composed from a multitude of disparate data sources) are presented to a human decision maker in a coherent way that provides critical context without overwhelming them with information. Users must be able to meaningfully interrogate the underlying data while also understanding the outputs of the AI tools that are augmenting that underlying data.

21. The system could also contain information on the effectors available to the soldiers in the field and utilise autonomy to suggest potential matches between target and effector. Here it is critical that all relevant information about the target/effector pairing be available to the decision maker, including availability, effectiveness, and potential collateral effects.

22. Ultimately, that information can then be presented to a human decision maker to take the decision on if the object is a legitimate target relevant to their current military priorities, and asses the proportionality of taking action with the available effectors (including collateral damage estimation and the presence of entities on a no-strike list). This final decision, about military necessity, proper use of scarce resources, and potential harm to non-combatants, is — and in our view, should remain — a human one. But it is potentially improved by thoughtful presentation of information and the output of automated processes.

23. This example is a far cry from a fully automated weapon system, but is closer to the reality of the situation we face today and are likely to face tomorrow. There may be space for fully autonomous weapons systems as one specific tool fit for very narrow operational circumstances, but any effort in that domain must be directed by a clear understanding of the realities of what is technically feasible and ethically defensible.

24. Full autonomy might be understandably appealing to some military decision makers worried about the speed and scale of targeting decisions that might to be made in a future war, but the example of autonomous vehicles offers an instructive example on why scepticism about the technical possibilities might be warranted. As far back as December 2015, Tesla CEO Elon Musk told Fortune Magazine that they had "all the pieces" to make self-driving cars work, theorising that it would require the efforts of thousands of people for two years.[1] In February of 2018, Musk claimed that fully autonomous capabilities would be available to customers within "six months, at the outside."[2] In a Jan 2021 earnings call Musk stated that "I'm highly confident the car will drive itself for the reliability in excess of a human this year… This is a very big deal."[3] The fact that we still do not have fully autonomous self-driving vehicles speaks to the complexity of the

[1] https://fortune.com/2015/12/21/elon-musk-interview/

[2] https://techcrunch.com/2018/02/07/elon-musk-expects-to-do-coast-to-coast-autonomous-tesla-drive-in-3-to-6-months/

[3] https://www.nytimes.com/2021/12/06/technology/tesla-autopilot-elon-musk.html

undertaking. Incremental improvements have surely been made (e.g., lane and parking assist) and some measure of autonomy may be feasible within the narrow confines of specific Operational Design Domains (ODDs), but even vehicle autonomy that works flawlessly under the vast majority of predictable conditions will likely never be good enough when errors in the marginal periods of operation can be fatal. Vehicle driving is complex by virtue of a full range of planar free movement, uncertainty in conditions, and continues flux in the surrounding environment, factors which only partially explain why vehicle automation has proved so difficult to reliably execute. It should therefore require no stretch of imagination to comprehend why fully autonomous manoeuvring in war is dramatically more complex with ever more catastrophic consequences for failures. That fully automated self-driving vehicles seem to have hit a plausibility barrier should give us pause on the prospects of full autonomy in a dramatically more complicated and consequential domain such as armed conflict.

25. However, the automotive example also speaks to the potential utility of AI and automation in constrained domains with specific and clear objectives that include human guidance and intervention. For example, features like warnings when a car strays outside of a lane or assistance in parallel parking have proven to offer tremendous utility and safety benefits. The fact that full automation remains a future dream does not mean that the potential for realising safety and quality improvements through sophisticated AI tools should be abandoned. But it does accentuate the need for clarity around the realities of the technology and the specific problems it can help address on the road. The same applies *mutatis mutandis* for AI in weapons systems.

26. We do not believe technology is yet capable of the kind and quality of choices that would have to be made for AWS outside of its current defensive applications and perhaps very narrow modes of offensive target engagement to be a reality, and that there may even be fundamental barriers to developing and deploying a generally applicable AWS that society would be willing to accept. There are so many edge cases and places of ambiguity in warfighting, where human judgement plays a critical role. If we go down the route of allowing for deterministic offensive decision making through an automated system, it would require — similar to the very limited cases of success with self-driving vehicles — a commitment to thoroughly defining and adhering to Operational Design Domains ("ODDs") for the application of specific AWS, where those domains are understood to be sufficiently resistant to uncertainties and ambiguities that constitute the fog of war. It would also mean acknowledging some level of residual uncertainty and a conscious willingness to accept a certain error rate with potentially grave or lethal consequences. That trade-off profile would need to be understood, justified by the military necessity of the situation, and consistent with the decision makers' obligations to uphold distinction and proportionality under IHL.

27. We are sceptical that there will be any generalisable AWS that meets those criteria. There may be some room for narrow, domain-specific AWS applied to certain target types, for instance AWS that is specifically designed to target tanks in certain weather conditions against certain terrain. But even in this narrow example, the number of edge cases and considerations that would need to be addressed speak to a rigorous test and evaluation framework, and the very real possibility that such a system should not be fielded even after years of development. If the system meets the high bar that society sets, commanders will need to thoroughly understand the domains where the AWS could be deployed, and ultimately assume accountability for every "decision" the AWS makes based on their decision to deploy.

28. We don't believe that there is a clear operational distinction between a world with AWS and a world without. Moreover, it is clear that autonomy and AI will become ever more important in warfighting applications. At the same time, the objectives these technologies seek to enable will need to broadened beyond current goals that may constitute the primary focus of technology development efforts. The militaries of the world typically are not – and should not be – looking to simply optimise on decision speed. Rather, objectives should be focussed on optimising decision speed and decision quality together. We therefore should be directing our efforts towards building tools that leverage the power of machines to enable the flexibility and creativity of human decision making necessary to adapt to the changing situations and priorities on the

battlefield, and avoid building rigid systems that, by virtue of over-reliance on machine automation, are in fact incapable of adapting to new situations and therefore may be prone to unexpected and even catastrophic failure when presented with novel situations.

**Question 3: What safeguards (technological, legal, procedural or otherwise) would be needed to ensure safe, reliable and accountable AWS?**

29. We would like to highlight two safeguards that we think are among the most critical: (1) ensuring that systems in this category are designed to "fail safe" to the extent possible, in both holistic system design and in User Interface/User Experience (UI/UX) design, and (2) investing heavily in training and education for the personnel that will be utilising these tools or deploying these systems.

30. Fail safe is a design principle that both expects failure as part of regular operations and gracefully handles that failure when it occurs. For instance, in the context of large scale, nation-state conflict, there is likely to be an incentive to make AI-related tooling, up to and potentially including AWS, responsive to any potential threats given the magnitude of those threats (e.g., a nuclear first strike). Functionally, this translates into system tuning in favor of false positives rather than false negatives. In this case, AI augmented human-in-the-loop systems can provide an effective safeguard against potentially catastrophic misapplication.

31. Additionally, one should consider fail safe measures such as manual override or seamless conversion to an operational mode with AI or other automation deactivated in response to, for example, malicious action, unaccounted adversary adaptation to AI, or identified failings. Each consequential aspect of the system should be designed to fail safe and considerable thought will need to be invested into defining those "consequential aspects" considering anticipated or intended deployment contexts and mission objectives.

32. As a general design principle, implementation of failing safely can guard against some of the worst-case potential end states of AWS. For example, for human-in-the-loop systems, fail safe measures can be further bolstered through interface features that present the output of AI tools along with supplementary information that helps to more fully contextualise that output for the end user. Providing users with an accessible interface that enables (or even requires) that they interrogate the underlying data and logic that lead to a significant analytical conclusion, confidence measures or other information to qualify the certainty of the conclusion, and oth er metadata alongside the specific output of the AI tool can lead to more informed, better operational decisions and outcomes.

33. While a great deal of responsibility will, and should, rest with the engineers building the systems, there must be a parallel effort to train personnel who will be using these systems. As the accountable human decision makers, those who are ultimately responsible for the critical decisions that will be made with these tools, they need to have an in-depth understanding of the systems they are looking to deploy. They will need to have some sense of the training data that was used to train the AI tools being used, the scenarios or domains where they will be more or less accurate, the appropriate purposes against which the tools can be deployed, experience using the tools in a variety of exercises and scenarios, and ultimately working knowledge of how to safely and effectively deploy these tools to augment their own decisions and fulfil their own mission responsibilities. Such comprehensive training is not just essential to help ensure the appropriate and responsible use of the systems, but is also an important step in enabling the end-users trust in the tools that they must rely upon to carry out their responsibilities.

**Question 4: Is existing International Humanitarian Law (IHL) sufficient to --ensure any AWS act safely and appropriately? What oversight or accountability measures are necessary to ensure compliance with IHL? If IHL is insufficient, what other mechanisms should be introduced to regulate AWS?**

34. IHL principles remain critical and should continue to guide both conduct in war and the development and deployment of any systems that include autonomy as part of a targeting process, including AWS. Though we believe IHL may provide a sufficient legal basis for underwriting the safe and appropriate use of AWS, it is increasingly clear that more work is required to translate these principles and obligations into real world mechanisms (e.g., practical guidelines, implementation standards, doctrine) that would bring sufficient oversight and accountability to AWS. Key requirements, such as the requirements for distinction, proportionality, and military necessity, must continue to guide both military decision makers and technologists looking to operate in this space. Ultimately, despite the suggestion of its name, Artificial Intelligence, including AI that informs AWS, is not an actual intelligence that can reason morally or be held accountable for immoral decisions.

35. Questions of distinction and proportionality ultimately require more than quantitative assessments, which may or may not be translatable into programmable code. They also require some significant measure of qualitative human judgement. Does one *reasonably believe* this target is a combatant? Is the potential collateral damage *acceptable* for the military objective in question? Have the relevant procedures been followed to ensure that the human has a *reasonable amount* of reliable information? These are all questions for which ranking commanders or oversight bodies could have a conversation with the human decision maker after the fact if necessary, and hold a specific individual accountable for mistakes.

36. Accountability in the case of an AWS would likely be an entirely different matter. The specific targeting determination or recommendation (e.g., "this set of pixels represents an enemy tank that should be targeted") may be logical or probabilistic inferences based on rule sets and tuning done to a standard specified in the handbook for the system. The system also might have, encoded in its program, some level of anticipated non-combatant damage above which it does not target the object. And of course, each of these "decisions" will have an anticipated error rate - this error rate might even be lower than a comparable human decision maker. But we will have moved from a world of specific, qualitative, morally weighted questions carried out through *procedurally-oriented* evaluations to aggregate, quantitative, optimised calculations with a largely *outcomes-based orientation.* This is not a change in degree, but rather a complete shift in kind with respect to methodologies and moral faculties for grappling with the challenges and dilemmas of warfare. It would require a new edifice of oversight and accountability mechanisms to ensure that distinction, proportionality, military necessity, and other IHL principles were implemented appropriately.

37. Again, we think it is critical to also focus on the IHL considerations of AI in weapons systems short of AWS. As a decision support tool, we should build AI that encourages thoughtful consideration of distinction and proportionality, that prioritises information that might speak to potential civilian harm, and encourage more informed human decision making throughout the military decision making process.

38. Finally, distinct from the *in bello* considerations of IHL, it is worth considering what *ad bellum* norms might need to be strengthened when it comes to strategic applications of AI and AWS that may play into nation-state decisions of when to go to war. The potential for unintended spirals is high when AI is deployed in tense situations that could escalate, where misunderstanding could lead to armed conflict. There should be a particularly high bar for the use of AI enabled systems in these contexts, and particularly human users should be empowered to exercise judgement to ensure that AI tools don't accidentally escalate a situation.

**Question 5: What are your views on the Government's AI Defence Strategy and the policy statement 'Ambitious, safe, responsible: our approach to the delivery of AI-enabled capability in Defence'? Are these sufficient in guiding the development and application of AWS? How does UK policy compare to that of other countries?**

39. In general terms, we support the Defence AI Strategy and its accompanying policy statement. We believe the former sets the right objectives and addresses the right underpinning fundamentals of

effective and responsible military AI – for example, the importance of testing and evaluation capabilities, sound data foundations, and a capable workforce. We also support the broad systems approach taken by the policy statement – i.e. that effective and responsible outcomes should be viewed as a function of norms and processes that exist across the Armed Forces and beyond (e.g., IHL), and not of any one instrument.

40. The challenge remains for the Armed Forces and wider UK Government to give practical effect to these policy directions, and to recognise both (a) the urgency with which the Armed Forces need to improve their AI capability, if the UK is to maintain its relative international position and (b) the sheer scale of the commitment required, not only by the Armed Forces, but across the defence-industrial base, the education system, and wider society.

41. One of our primary concerns relates to the Armed Forces' ability to access AI talent – that is, those with a background in fields such as computer science and data science, and underlying disciplines such as mathematics and physics (when thinking of AI's skills requirements, it may be helpful to think of AI as "advanced maths"). The UK is fortunate to have deep and in some respects world-leading reserves of such talent. And yet the Armed Forces struggles to access it. Sharply uncompetitive remuneration is one reason for this, but there are others, including inflexible MOD career structures (across civilian, uniformed and reserve workforces) and a wider lack of recognition of how those with AI skills can deploy their talent in service of a critical national mission.

42. While expertise can be contracted in – and is, at significant cost – there is no substitute for the Armed Forces having their own internal AI expertise. Among other things, this is needed to ensure that the Armed Forces are informed and empowered as they set strategy, commission capabilities, hold suppliers accountable, and otherwise engage with private sector actors (the driving force of most AI developments).

**Question 6: Are existing legal provisions and regulations which seek to regulate AI and weapons systems sufficient to govern the use of AWS? If not, what reforms are needed nationally and internationally; and what are the barriers to making those reforms?**

43. As we note above, we believe that the existing tenants of International Humanitarian Law provide a critical and enduring core for AI and weapons systems regulation, but that they must also be further articulated, strengthened, and perhaps extended to govern the use of AI in weapons systems. Principles like distinction, proportionality, and military necessity will need to be implemented in regulations and procedures that govern the development and deployment of AI in weapons systems, with particular attention to testing and evaluation, ongoing maintenance and reliability, domains of applicability, and interpretability.

44. While existing IHL provides a firm basis in terms of core legal provisions and regulations that may be further adapted to regulate AI and the use of AWS, we believe that specific attributes of AI technologies may lend themselves to a particular suggestion for reform. We suggest consideration of provisions to address the inherent brittleness of AI technologies that is often ignored, misunderstood, or even purposefully misrepresented for commercial, programmatic, or publicity benefits and to the extreme detriment of the institutions reliant on these technologies.

45. For example, Machine Learning (ML) techniques (as one prominent category of AI technology) often produce the suggestion to lay observers — by virtue of the very title — that these systems are actually continuously and in real-time "learning" and "adapting" to novel situations and parameters similar to humans and in ways that make them resilient to ever-changing application environments. The reality of the most implementations of these technologies is, however, quite far from that suggestion. In actual standard practice, ML describes the technique for *building* (rather than *operating*) models that, once trained, lock in a set of parameters that are to remain fixed until the model is updated based on new data, model features, optimization parameters, etc. This process of version control is a function of the fact the ML models, like other technologies, should in strict terms be expected to perform with some degree of predictability based on tested

and validated specifications. It is a requirement, in fact, a requirement for addressing the inherent brittleness — i.e., the tendency of models be anti-resilient to changes in data, environment, application context, etc. — of ML models.

46. And yet a prevailing assumption — based on misconceptions of AI as something more akin to human learning or intelligence — is that these technologies are truly resilient, adaptable, and transferable to virtually any new environment.

47. Formal regulation on AI, especially as applied in AWS, that focusses on addressing inherent brittleness, and provides assurance regarding the performance of models, would help close the gap on at least three areas of deficiency that we see as imperilling the effective and responsible use of AI technologies:

    1. ***Capability Clarity*** *— It would compel clear articulations of operational domain conditions, constraints, and limitations for AI technologies.*

    2. ***Deployment Durability*** *— It would emphasise technology researcher, manufacturer, and program developer responsibilities for not just short-term testing and early deployment outcomes, but also long-term maintenance assurances and product liabilities for the sustained delivery of marketed results.*

    3. ***Hyperbole Humbling*** *— It would require AI capabilities developers and providers to curtail hype-driven impulses by minimizing the use of confusing, distorting, or misguiding language (e.g., anthropomorphising phrases and characterizations) and generally incentivising the representation of capabilities grounded in truthful, verifiable, and reproducible claims.*

48. We do not wish to be prescriptive on specific implementations of this suggestion, but for the purposes of illustrating a potential legal hook, we suggest that this requirement to address inherent brittleness might be well-situated as an amended critical component of Article 36 weapons reviews.