Duality Technologies Response to Request for Information (RFI) Related to NIST's Assignments Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial Intelligence (Sections 4.1, 4.5, and 11)

Duality Technologies, Inc., thanks the National Institute of Standards and Technology for the opportunity to contribute the following comments in its response to the Request for Information Related to NIST's Assignments Under Sections 4.1, 4.5, and 11 of the Executive Order Concerning Artificial Intelligence.

Duality[1] is a leading Privacy Enhancing Technology (PETs) provider, enabling organizations to collaborate on sensitive analytics and AI while utilizing personal data and other sensitive data. Our Platform offers a broad set of privacy technologies – including hardware and software solutions – to enable AI models to be deployed while protecting privacy and security. The Platform offers governance tools to manage how these privacy technologies are used – i.e., roles and access controls, schema management, logging, etc. This combination of capabilities, along with data science and analytical models, allow organizations to collaborate with one another on their most sensitive data in order to derive new insights while ensuring compliance with regulatory requirements and their own business policies.

Our software platform leverages several PETs including Fully Homomorphic Encryption (FHE), Secure Multiparty Computation (SMPC), Federated Learning (FL) and Federated Analytics (FA), and Trusted Execution Environments (TEE). For FHE, we utilize OpenFHE - an open-source, standards-compliant Homomorphic Encryption library built in part with US Government funding[2]. Duality's founding team is comprised of world-renowned cryptographers, including Turing Award winner Prof. Shafi Goldwasser, Gödel Prize winner Prof. Vinod Vaikuntanathan, DARPA Fellow and co-founder of the HomomorphicEncryption.Org standards body Dr. Kurt Rohloff, and data science experts.

NIST's AI RMF rightly points out that AI has "significant potential to transform society and people's lives", yet also poses numerous risks to individuals, organizations, and more. Given our grounding in privacy technologies, our feedback and comments touch on the following areas from a data privacy and data protection perspective:

- Risks and harms of AI and Generative AI
- Changes that users and organizations can make to mitigate AI risks
- The types of technologies and governance methods that can mitigate these risks
- Best practices for implementing and leveraging these technologies
- Suggestions for application specific standards that AI Actors would benefit from

Overall, our feedback is focused on best practices related to the use of sensitive data and analytics in an AI context. We suggest measures to mitigate risks while maximizing outcomes, particularly by leveraging proven privacy technologies and technically enforced governance.

Section 1: Duality Responses to Questions around "Developing Guidelines, Standards, and Best Practices for AI Safety and Security"

Recommended changes for AI Actors to make to their current governance practices to manage the risks of generative AI

[1] https://dualitytech.com/
[2] https://www.openfhe.org/community/

The wave of growth and innovation brought on by AI is undeniable, but it's true potential can only be unleashed when models can be trained and deployed on the best existing data – even if it's not necessarily readily available to specific users and teams in a given organization, or to the general public. This indeed includes utilizing sensitive data that might be spread out across a number of siloes - e.g., across different organizations - making accessing and using it fraught with security, privacy, and trust challenges.

This has immediate economic impacts to the growth of AI, and to the value it can yield for its users. Enterprises are already running into difficulties in leveraging AI for this very reason, which impacts both AI Providers and Users alike. Organizations that hold data will not share that data with Providers because of data protection concerns, while AI providers do not want to share their intellectual property without appropriate protections. As a result, and if the current state continues, AI will not deliver the outcomes it promises.

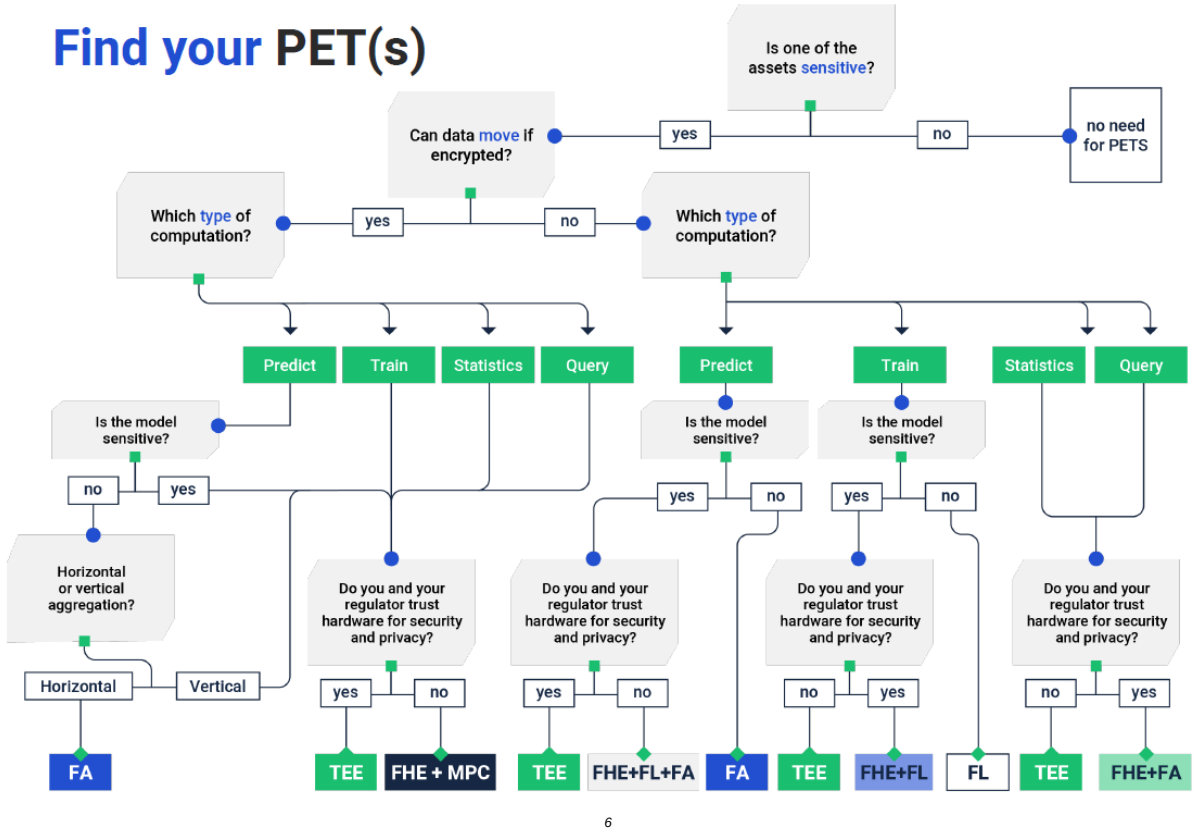
Managing the risks of generative AI means recognizing them. From our perspective as a PETs provider, and in our experiences in allowing our customers to utilize sensitive AI models and data, these risks include privacy, security, and access to appropriate data. Mitigating such risks has a positive economic impact - enabling growth of markets around AI models and data, while at the same time reducing privacy, security, compliance, and cultural blockers.

Given the above, changes that AI Actors should consider implementing include:

- Incorporating privacy as a part of the standard Machine Learning Operations (MLOps) pipeline. This means adding PETs into their operational pipeline for training and tuning models as well as for serving them. Practically, this entails leveraging tools that protect sensitive data while both building and customizing models and deploying them in an operational cluster.
- Leveraging a combination of process and technology-based solutions to ensure privacy.
    - To level-set, Privacy enhancing technologies are technologies that support protection of data-in-use - including data security, data privacy, and confidentiality. The *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* references both PETs and Privacy Preserving Technologies (PPTs) – terms which can be used interchangeably. For the purposes of this response, we will use the commonly accepted definition from The European Union Agency for Cybersecurity (ENISA), which defines these as "software and hardware solutions, i.e. systems encompassing technical processes, methods, or knowledge to achieve specific privacy or data protection functionality or to protect against risks of privacy of an individual or a group of natural persons"[3]. Per the UK's Information Commissioner's Office, PETs support "data protection by design", and can help organizations "implement…data protection principles effectively and integrate necessary safeguards

[3] European Union Agency For Network And Information Security (ENISA), Marit Hansen, Jaap-Henk Hoepman, Meiko Jensen, *Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies Methodology, Pilot Assessment, and Continuity Plan*, 2015, https://www.enisa.europa.eu/publications/pets/view/++widget++form.widgets.fullReport/@@download/Readiness+Analysis+for+the+Adoption+and+Evolution+of+Privacy+Enhancing+Technologies.pdf

into…processing"[4] and in fact recommend their use[5]. The ultimate benefit of PETs is that they enable users to leverage almost any data - including sensitive personal and business data - without compromising on privacy and security.
    - There are many different types of PETs which can be used alone or in combination to support various privacy-protected data processing needs. The chart below shows some examples.
- It is also important to ensure the technologies referenced above and throughout this response are connected to appropriate governance and data management tools. This should include managing roles and permissions around access to certain data throughout the data science pipeline, restrictions on who can initiate which types of analyses, and who can see which results.



6

Chart designed by Duality demonstrating where different PETs are best fit for purpose.

These risks manifest themselves in different ways, and all prevent the extraction of insights and/or economic benefit. To help further illuminate the need for privacy in order to manage the

[4] https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/privacy-enhancing-technologies/how-can-pets-help-with-data-protection-compliance/
[5] https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2023/06/ico-urges-organisations-to-harness-the-power-of-data-safely-by-using-privacy-enhancing-technologies/
[6] Please find definitions of each PET in the attached appendix

risks of AI broadly speaking, and generative AI in specific, below are some examples of use cases where these challenges arise:

1. Public Sector: Collaborative Model Training: How can government agencies across the public sector work together, and potentially even include private sector organizations, to build and deploy a variety of models? These models could include cybercrime or financial crime detection models, public health models, national security models, etc.
    a. Privacy is needed in these cases both because of the sensitivity and sources of the data, as well as the need to protect the models themselves. In fact, DARPA has funded research around this challenge, for example in the DARPA Cooperative Secured Learning (CSL) Project[7], in which Duality participated.[8]
2. Healthcare - Predicting Pathologies: How can organizations deploy AI models to predict and detect pathologies using imagery data linked with Personally Identifiable Information (PII) and Personal Health Information (PHI), while also complying with HIPAA?
    a. Privacy is needed in this situation in order to comply with the law and enable more in-depth utilization of data at higher levels of accuracy and precision than is typically possible (since organizations often opt to simply delete sensitive fields to maintain HIPAA compliance, generalize them, randomize them, or even replace them with synthetic data, which negatively affects results)[9]
3. Financial Services - Risk Scoring: How can organizations build better risk models by combining features across data provided by vendors and financial institutions, while ensuring sensitive data and models are protected?
    a. Privacy is needed in this situation in order to protect intellectual property and mitigate economic risk. Depending on the data, privacy may also be needed to protect personally identifiable information and to comply with the law
4. Health Risk Prediction and Precision Medicine: How can genomic data and other sensitive PII and PHI be linked, and then models be deployed to predict health risks and enable precision medicines, for example on certain types of cancers
    a. Privacy is needed for regulatory compliance as well as the ability to compute on sensitive personal health information at levels of accuracy and precision not possible with typical methods, which are mentioned above. NIH has funded research around this, which Duality supported[10] and then built on further[11].
5. Accelerating Model Evaluation: How can organizations test third party models and generative AI offerings on real data before proceeding with a purchase and productization decision.
    a. Privacy is needed because these processes are often significantly hampered by IP and data sensitivity concerns.

Economic and security implications for watermarking, provenance tracking, and other content authentication tools

As demonstrated above, inserting privacy and confidentiality into AI-adoption processes is crucial to the economic success of this technological category. Privacy technologies allow for

[7] https://www.darpa.mil/program/cooperative-secure-learning
[8] https://www.prnewswire.com/il/news-releases/darpa-contracts-with-duality-technologies-to-develop-privacy-preserving-machine-learning-for-covid-19-research-301096126.html
[9] https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html
[10] https://pubmed.ncbi.nlm.nih.gov/32398369/
[11] https://www.pnas.org/doi/abs/10.1073/pnas.2304415120?doi=10.1073/pnas.2304415120

intellectual property to be protected, which also of course has national security advantages as well.

In the world of AI, models certainly qualify as intellectual property. With certain types of AI, like Generative AI, a model is also a reflection of the data that was used to train it, meaning that in some cases, protecting the model is also critical to protecting the underlying data. An example of this came to light in late 2023, when researchers found that an LLM model would reveal its underlying training data - including personal information - when asked to repeat a word several times.[12]

The methods mentioned by NIST, like watermarking, provenance tracking, and other content authentication tools certainly contribute to protecting IP in this context, but there are other tools worthy of mention. One specific hardware-based tool is a "Trusted Execution Environment" (TEE). A TEE is a secure hardware environment where AI models can be trained, tuned, and deployed, however an outsider would only be able to see encrypted ciphertext. TEEs are one of the best known methods for protecting model and data IP in the world of AI today (others, like FHE, are closely following), and the implications of using this technology mean that models can be trained and run against the best *existing* data, not just the data an organization happens to own, and also made available for more people and organizations to use.

The need for greater controls when data are aggregated

Aggregating data creates the potential for significant risk when using both sensitive and ostensibly non-sensitive data. It is indeed important to highlight that even seemingly non-sensitive data can become very sensitive when aggregated. An example of this could be aggregating anonymized social media posts with location data. One could potentially cross-reference this data to reveal sensitive information about an individual's whereabouts, daily routines, home address, etc. Another example could be combining public transportation schedules with employee attendance records - individually these data sets may be non-sensitive, but together could reveal movement patterns, which could have safety and security implications.

As such, aggregation in and of itself creates privacy and security risk, even when seemingly innocuous data is used, creating targets for nation states adversaries, organized crime, and more - not to mention also being vulnerable to insider threats.

Additionally, once data is aggregated, even if it remains non-sensitive, certain queries or models run against that data - and the insights those generate - could be very sensitive. This calls for not only protecting the data set itself, but also the analysis used, as well as the result.

Using technologies that are generally available today, it is possible to analyze distributed databases, where the data used for AI training, tuning, and computations can be housed in different locations, and potentially controlled by different data owners, yet authorized users can still run computations against the data holistically, as if it were a single data set, in a privacy protected manner. Leveraging secure and privacy protecting technologies like FL and FHE together can enable this.

[12] https://www.cpomagazine.com/cyber-security/security-researchers-chatgpt-vulnerability-allows-training-data-to-be-accessed-by-telling-chatbot-to-endlessly-repeat-a-word/

Within this decentralized topology, it is also possible to protect the models themselves, which may be needed if the model contains business or government-sensitive information – for example, how certain models work, or what they are looking for, or what results they generate. Again, this architecture can be supported by PETs.

Creating guidance and benchmarks for evaluating and auditing AI capabilities - Availability and gap analysis of methods for measuring AI systems' privacy

Using third-party AI services, e.g., summarization tools, can compromise privacy in a number of ways. For example, the AI Vendor could be exposed to sensitive prompts, or even their customers' data used for model training. Mitigating these risks is particularly important if the AI service utilizes its own internal / proprietary model, rather than relying on a more open provider, such as OpenAI, or other established foundation models. To mitigate this risk, it is important for users to understand the service's architecture. Specifically, users should determine whether the service uses its own model or integrates external AI services. As a best practice, users should confirm that the provider uses a combination of PETs and governance tools to enforce safeguards, and that prompts and training data are always encrypted and hidden from the AI Vendor, particularly if one or both are sensitive.

Section 2: Duality Responses to Questions around "Advance Responsible Global Technical Standards for AI Development"

Best practices regarding data capture, processing, protection, quality, privacy, transparency, confidentiality, handling, and analysis, as well as inclusivity, fairness, accountability, and representativeness (including non-discrimination, representation of lower resourced languages, and the need for data to reflect freedom of expression) in the collection and use of data

Duality can comment on best practices regarding data capture, processing, protection, privacy, and confidentiality. Privacy, security, and confidentiality are essential to ensuring values such as non-discrimination, eliminating bias, and freedom of expression are upheld. In fact, this is the driver behind one of the first implementations of Fully Homomorphic Encryption, a PET, are a series of projects sponsored by DARPA culminating in the OpenFHE open-source Fully Homomorphic Encryption software library. These projects, over multiple years, have led to hardware acceleration of FHE[13], its application for software obfuscation and machine learning, and adoption to secure 5G against nation-state adversaries[14].

When it comes to AI development, maintaining these protections and upholding these values are both necessary. The reality is that any one organization or actor cannot be expected to have the best, fit-for-purpose data for any given analysis they would like to run, or model they would like to develop. On the contrary - the best data is often spread out across multiple actors, and only by accessing and leveraging it could one ensure an optimal outcome from an AI modeling and insight generation perspective, while also reducing challenges around data quality, representation, and fairness.

These activities are often blocked because of data sensitivity (i.e., data that could be used to build a model with biases would likely include sensitive aspects encompassing personal or business-sensitive data), or even regulation (e.g., data sovereignty). Therefore, in order to

[13] https://www.prnewswire.com/news-releases/darpa-awards-duality-technologies-multimillion-dollar-contract-to-accelerate-machine-learning-on-encrypted-data-301717233.html
[14] https://www.darpa.mil/program/open-programmable-secure-5g

actually utilize the best possible data for AI purposes, Actors must leverage solutions that incorporate privacy, confidentiality, and data protection throughout the entire analytics lifecycle, from data capture, to model development and training, to inference, etc.

The best practices for doing so today involve the use of PETs, combined with appropriate governance. These technologies and platforms facilitate access to the best possible data, whether via a federated approach where data does not move at all, or a centralized one, while protecting privacy.

Best practices for AI model training

Again, we can comment on this from the perspective of protecting data privacy, model privacy, and security. The main challenges regarding AI model training when sensitive data and/or models are used are:

- When leveraging sensitive data, not every party must be exposed to it in the training phase. This is especially true when multiple parties come together to train a model (whether federated or aggregated)
- If the intermediate results are shared amongst collaborating parties, information about the underlying data can also be inferred - in some situations you want to protect against this, as with the DARPA CSL project referenced earlier
- Sometimes an AI Actor cannot aggregate data in order to use it for training (e.g., in situations of very sensitive data, data sovereignty across countries, etc.)
- It is often not possible to conduct Exploratory Data Analysis on sensitive data
    - When creating AI models, data scientists must first know the schema and explore the data itself, with the goal of deciding on the required feature engineering and optimal model type. If the data is sensitive and the data scientist is unable to use real data, he or she will often turn to sample data (which by definition is incomplete) or synthetic data (which comes with its own set of challenges around precision and accuracy).
    - Ultimately, the data scientist must decide if a model works well or not, and fine-tune it, and this is optimally done on real, complete datasets

As such, best practices we often see include:

- Leveraging PETs - specifically TEE, FHE, FL, or a combination thereof (e.g., FHE and FL). These PETs can help protect data and model privacy, and can enable various AI model training modalities, including in situations where data cannot move
- The AI Actor must determine what risks he or she is trying to protect against, which laws must be complied with, etc. Below is an example of what types of questions must be considered when analyzing data (whether using AI or not), and mapping those to applicable technologies to support risk mitigation (see previous diagram, titled “Find Your PET”).
- When leveraging PETs or other risk mitigating technologies, these must be integrated with appropriate technical and organizational governance. Defining specific best practices around these would depend on various legal, regulatory, and security requirements and could be an area where NIST’s input would be impactful
- If using cryptographic-based PETs, like FHE, ensure the use of an open-source and standardized library like OpenFHE; if utilizing hardware-based PETs, ensure there is the appropriate level assurance on the hardware itself
- After models are trained, it is often important for collaboration parties to pre-define specific analysis and what types of results are permitted in order to ensure the adequate

level of data protection (since asking certain questions / doing certain analyses might reveal sensitive data). It is important that PET platforms have such capabilities around governance and permissions of queries, data, and results to enforce these agreed-upon computations.

- Concretely, in the near term, TEEs are often the optimal privacy technology to leverage because they allow the widest breadth of data and models to be utilized. However, in the longer term, FHE is preferred for the following reasons (note: this is a non-exhaustive list):
    - Portability - FHE is a software-based solution and can run on any hardware, while TEEs are hardware based, meaning the use of specific hardware is needed
    - Trust - FHE is based on cryptographic trust, which can easily be verified based on open source libraries and checked for compliance against existing standards. TEEs inherently depend on the security of the underlying hardware and firmware
    - Strength of encryption - Software-based encryption is considered to be stronger than hardware-based encryption
    - Susceptibility to attacks - FHE is more resilient against side-channel attacks, while the use of hardware increases risk
    - Quantum Resistance - FHE is considered to be resistant to attacks from quantum computing devices, whereas TEEs rely on traditional cryptographic algorithms that could be more vulnerable to these types of attacks
    - Performance - performance of FHE is improving by orders of magnitudes, and as such increasingly complex models can be trained or executed on homomorphically encrypted data.

Again, we appreciate the opportunity to provide feedback on NIST's request for information, and to share more about how privacy technologies can support national, commercial, and individual goals in the use of Artificial Intelligence.

Appendix 1: Definition of PETs shown in “Find your PETs” chart

- Fully Homomorphic Encryption (FHE): A post-quantum encryption scheme that enables computations to be run on encrypted data, and/or to deploy encrypted models. The model/data remains encrypted throughout the entire computation process, yielding encrypted results that can only be decrypted by the permissions party.

- Trusted execution Environment (TEE): A secure and isolated area within a server that ensures the confidentiality and integrity of data and processes that are executed inside it. TEEs allow a pre-approved set of computations and prevent access to raw data by any external party. Most often used to train and deploy AI Models, including generative AI.

- Federated Analytics (FA): A privacy-preserving framework for computing data analytics over multiple remote parties (e.g., mobile devices) or siloed institutional entities. With FA, the computation is performed locally and the data never leaves the premises.

- Federated Learning (FL): A subset of FA that describes a machine learning technique that trains models on multiple remote datasets. With FL, the training is performed locally and the data never leaves the premises (i.e., in a decentralized manner, without moving any data).

- Multi-Party Computation (MPC): A subfield of cryptography with the goal of creating methods for parties to jointly compute a function over their inputs while keeping those inputs private. With MPC, each party only receives part of the data, thus no single user can access the data itself. Typically used in conjunction with FHE in order to allow multiple parties to collaborate.