Comments of the Software and Information Industry Association

Request for Information: Executive Branch Agency Handling of Commercially Available Information Containing Personally Identifiable Information

Office of Management and Budget

December 16, 2024

The Software & Information Industry Association (SIIA) appreciates the opportunity to provide comments on the Office of Management and Budget's (OMB's) request for information regarding executive agencies' handling of commercially available information (CAI). SIIA is the principal trade association for companies in the business of information, including its aggregation, dissemination, and productive use. Our members include roughly 375 companies reflecting the broad and diverse landscape of digital content providers and users in academic publishing, education technology, and financial information, along with creators of software and platforms used by millions worldwide, and companies specializing in data analytics and information services.

We recognize the important goals of this inquiry and appreciate the thoughtful approach taken by OMB to the questions. Our goal in these comments is to provide OMB with recommendations on how the federal government should approach handling of CAI containing PII that agencies acquire from third-party data and analytics companies for non-national security purposes in a responsible manner.

General Considerations (Questions 1-3)

Federal agencies use CAI for numerous legitimate government purposes, including anti-fraud efforts, protecting consumer health and safety, counter-terrorism functions, and disaster relief efforts.[1] Agencies use CAI subject to a robust legal governance regime. When the CAI contains personally identifiable information (PII), agencies must comply with the requirements of the Privacy Act of 1974, the Federal Information Security Modernization Act, OMB guidance including Circular No. A-130, the Fourth Amendment of the U.S. Constitution, and other authorities. This tapestry of legal authority applies regardless of an agency's use of AI tools to make sense of the data it has acquired either directly from the public or through third-party data vendors. While we do not purport to provide the government with guidance about when and how CAI can assist in its mission, we do recommend a presumption that using CAI for legitimate governmental purposes is permissible so long as it abides by the Constitution and other applicable legal authorities.

The principal risks to individuals' privacy stemming from the use of CAI containing PII are the possibility of inadvertent disclosure of PII through cybersecurity breaches; mishandling of datasets containing PII; and the inappropriate use of PII to render legal or similarly significant decisions that have an adverse effect on an individual's access to government benefits or legal rights. In some cases, such as

[1] See also, e.g., ODNI Senior Advisory Group Panel on CAI, Report to the Director of National Intelligence (Jan. 27, 2022), at 7-10 (providing examples of how CAI contributes to the government's intelligence mission).

in the law enforcement context, inappropriate use of PII could raise questions under the Fourth Amendment. Guardrails against mishandling or inappropriate use are contained already in the legal tapestry that governs agency use of data.

What AI adds to this mix is the potential to generate greater analytical insight relevant to governmental functions. To the extent that this analytical capacity augments potential risks to individual privacy, we strongly recommend that the government focus on building targeted guardrails for high-risk AI systems – which in this context we define generally as rights-impacting AI systems[2] – rather than creating new limitations on the government's acquisition of CAI. So long as the government's use of CAI is legal, we do not see any need to treat AI that relies on CAI containing PII any differently than AI that relies on other datasets that contain PII.[3] The cost of imposing additional restrictions on the government's ability to use AI technologies to advance these types of initiatives could be significant and have unintended consequences that need to be carefully evaluated ahead of time.

In addition, we strongly recommend that any such guardrails be tailored to the ways in which those AI systems are used. Automated decisionmaking tools that determine individuals' eligibility for government benefits (such as eligibility for the Supplemental Nutritional Assistance Program (SNAP) or federal student aid) require more robust guardrails than do AI tools deployed to detect indicia of money-laundering or human trafficking activity by a financial sector enforcement agency. While the AI systems must be reliable and trustworthy in each scenario, the sensitivity of decisionmaking around individual rights and benefits requires an extra degree of human oversight.

As discussed below, OMB may wish to explore ways to encourage, if not require, responsible data practices by all agencies that are tailored to the type of data involved and the degree of AI processing of that data. As just one example, this could include additional cybersecurity requirements for AI training data that includes highly sensitive PII data being processed by automated decisionmaking tools.

Definition of CAI (Question 4)

The EO 14110 definition of CAI is likely to generate confusion because it defines CAI more broadly than the term is understood in most contexts. EO 14110 defines CAI as "any information or data about an individual or group of individuals, including an individual's or group of individuals' device or location, that is made available or obtainable and sold, leased, or licensed to the general public or to governmental or non-governmental entities."[4] This definition presumes that CAI will contain data about individuals. In fact, CAI is generally defined to include any information or data – not limited to information or data about individuals – that is available to the public on the open market (not to include

[2] See OMB Memorandum 24-10 (Mar. 28, 2024), section 5.

[3] We do not express an opinion about the impact of the Supreme Court's decision in *Carpenter v. United States* (2018) on federal agencies' acquisition and use of CAI.

[4] EO 14110, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Oct. 30, 2023), section 3(f).

information collected via classified or national security channels). This broader definition of CAI is reflected, among other places, in the recently issued Intelligence Community Standard 206-1.[5]

A narrow definition of CAI that presumes CAI includes PII can have unintended consequences. These may include policies and regulations that restrict the ability of the government to acquire and use CAI for purposes that have no direct bearing on individuals' access to benefits or legal rights even if data may contain PII. It is important that any new policies be developed with appropriate regard for how different agencies use CAI, including methods to protect PII if applicable, and the operational needs of those agencies.

Transparency Into Agency Handling of CAI Containing PII (Questions 5-8)

We appreciate the policy interests around improving government transparency around how agencies handle CAI containing PII. Nevertheless, except as noted below, we do not recommend imposing additional requirements on agency use of CAI containing PII beyond what is already required under law. We have concerns that doing so would impose significant burdens on agencies and not lead to any meaningful benefits to U.S. persons.

We do recommend that agencies adopt procedures to ensure that individuals are made aware of the basis of information that leads to a consequential decision – one affecting an individual's key rights or access to benefits.[6] Outside of the national security context, individuals should understand the information relied on to reach a decision about access to government benefits. For example, were an agency to deny access to SNAP benefits because an individual, as suggested by third-party data, visited an abortion provider, that would be important to know and may give the individual a basis to seek legal redress. However, in supporting transparency measures of this sort, we recommend that disclosure of all sources of relevant information, not limited to CAI. Some of these procedures may already be in place particularly when decisions are conveyed through formal processes (such as decisions on veterans' medical claims).

Agency Processes for Responsible Handling of CAI Containing PII (Questions 9-12)

In this section we provide five recommendations to mitigate risks associated with maintaining and using databases containing PII to drive AI applications and analysis.

First, we recommend that OMB approach privacy risks that may be exacerbated by AI tools by applying a risk-based framework. This includes focusing on high-risk applications of AI systems that rely on PII. The degree of risk to consumers from agencies' use of AI may have less to do with the data itself, or even the extent of processing by AI tools, than the sensitivity of agency *decisionmaking* that relies on this data. After all, ostensibly "sensitive" data may be leveraged for routine or inoffensive purposes. Meanwhile, even public domain information may be used to make critical decisions around, for example, government benefits. As such, we recommend that OMB guidance focus on high-risk AI uses, such as decisions about benefits and affecting legal rights. We believe these measures are reflected in existing

[5] Intelligence Community Standard 206-01, at Appendix A: Definitions.

[6] *See, e.g.,* C.R.S. § 6-1-1701(3).

OMB guidance around agency use of AI, such as Memoranda 24-10 and 24-18 and do not require additional safeguards.

Second, we recommend that OMB direct agencies to implement best practices in cybersecurity to mitigate risks associated with PII in agency-maintained datasets that are used to train AI systems. This may include conducting reviews of existing cybersecurity and cloud vendors.

Third, we recommend that OMB direct agencies accelerate their adoption of privacy enhancing technologies (PETs) to protect against misuse or unintentional disclosure of PII. This recommendation is not intended to be limited to CAI and should apply across the board to all datasets containing PII. For AI use cases, PETs allow users to securely train and evaluate ML models using data sources across silos and boundaries, including cross-jurisdictional, third-party, and publicly available datasets. By protecting AI models and workflows during processing – i.e., "data in use" – PETs can enable secure AI capabilities that enhance scientific research, discovery, and implementation. In addition to enabling net-new data usage, PETs also help ensure sensitive assets, including ML models trained over regulated data sources, remain protected at all points in the processing lifecycle. PETs eliminate the need to replicate or pool data, allowing data owners to retain positive control of their assets and limit the risk of misuse and unintended exposure. These solutions give users flexibility, enabling data value extraction while protecting the interests of all stakeholders. PETs allow government actors to leverage models trained on sensitive data over new datasets that might otherwise be unusable. The IC Policy Framework for CAI recommends the use of PETs, and we recommend OMB encourage other agencies to adopt PETs as well.

Fourth, we recommend that OMB provide agencies with guidance to distinguish reputable third-party data vendors from those without strong data stewardship practices. This, too, is reflected in the IC Community Policy Framework for CAI,[7] which directs IC elements to "assess the integrity and quality of CAI they access or collect," among other things. As discussed below, SIIA is currently working with industry to develop a series of data stewardship principles to guide responsible parties in the third-party data and analytics industry. We anticipate these can be relied on by the government to assess the integrity of the CAI that it acquires to assist in undertaking governmental functions.

Fifth, in situations where the government is undertaking functions that could have a direct impact on Constitutional rights, additional guardrails may be required. We recommend that the DOJ Office of Legal Counsel provide guidance as needed to ensure comportment with the law. In the law enforcement context, for example, unique safeguards may be needed to strike the right balance between privacy and civil liberties and operational needs.[8]

Other Considerations (Questions 13-14)

As a trade association working with companies across the information landscape, SIIA has a deep understanding of the ways in which governments use third-party data to address critical, societally beneficial needs. The positive use cases of data are often clouded by harmful data broker practices.

---

[7] Intelligence Community Policy Framework for Commercially Available Information (May 2024).

[8] The IC adopts this approach in the May 2024 Policy Framework. That approach could serve as a template for other sensitive uses of CAI containing PII, recognizing that the guardrails around IC usage must be more robust than in other contexts because of the legal restrictions on intelligence activities involving U.S. persons.

These practices have led federal and state policymakers to explore regulatory hurdles for government acquisition of third-party data that may have unintended consequences on legitimate governmental functions and on the Americans who rely on them. With this background, SIIA has been working closely with good faith data and analytics companies to develop data stewardship principles to guide data and analytics companies in responsible practices. We expect to publish these principles in early 2025 and we hope they will prove a useful resource in assisting federal agencies that engage with third-party vendors.

* * *

Thank you for considering our feedback on this RFI. We are happy to discuss any of these comments in further detail. SIIA's points of contacts for this submission are Paul Lekas, Senior Vice President for Global Data Policy (plekas@siia.net) and Anton van Seventer, Counsel for Privacy and Data Policy (avanseventer@siia.net).