The Sedona Conference WG11 Brainstorming Group Outline – Artificial Intelligence (April 2024)

The Sedona Conference WG11 Brainstorming Group Outline – Artificial Intelligence (April 2024)

Brainstorming Group Members:

David Sella-Villa (Brainstorming Group Leader)
Julian Ackert
Mark Allen
Hon. Ralph Artigliere (ret.)
Hon. Laurel Beeler
William Belt
Melissa Dalziel
Xavier Diokno
Ashley Picker Dubin
Gail Gottehrer
Cara Elizabeth Green
Anya Korolyov
Jon Polenberg
Marcy Stevens
James Sherer
Jake Simpson
Doug Meal (Steering Committee Liaison)

To: Members of WG11

From: WG11 Steering Committee

As reflected in the attached outline, the AI Brainstorming Group has been doing a lot of brainstorming but has not yet come to a consensus on what they recommend regarding the legal issue(s) that are worth addressing via a Commentary and how those issues should be addressed so as to move the law forward. We are hoping to get some feedback from you on those questions at the upcoming meeting. On behalf of the brainstorming group, we thank you for your help.

Sedona Conference Working Group 11
Artificial Intelligence Brainstorming Group
Draft Outline

Introduction

The technological development of Artificial Intelligence (AI) has reached a critical moment relative to the law. Today, enough AI applications are accessible and have entered the stream of commerce that almost any organization, and the professionals serving them, can use AI to further their work. This presents countless opportunities, the benefits of which are hard to ignore. The challenges, however, are harder to describe. This is particularly the case in the fields of privacy and cybersecurity.

Any new technology presents privacy and cybersecurity risks. The rapid development and adoption of AI has made it difficult, even for the most experienced in these fields, to find certainty in both defining the issues and tailoring the solutions. Practitioners in the fields of privacy and cybersecurity are often asked to follow the data. When it comes from or is shared with individuals, consent and disclosure often arise. Using the data for certain kinds of decision-making (e.g., in litigation, by public bodies, for profiling, etc.) triggers other privacy and security obligations.

This Working Group 11 Brainstorming Group explored the interaction between data privacy laws and AI. Because both are moving targets, the data privacy laws in questions were those in effect prior to 2023. These laws were promulgated when AI, and particularly generative AI, was not generally available. If there were questions about how the privacy laws applied to AI systems, only those organizations large and sophisticated enough to develop their own systems had to think through those problems in their respective instances. In short, the pre-2023 data privacy laws did not contemplate a world where AI was ubiquitous.

Since 2023, both changes to the existing privacy laws and new AI-specific laws have been proposed and acted. All of these new rules contemplate a new technological reality of AI being available to almost anyone with an internet connection. None of these new laws seem to be intended to override or reduce the obligations under the pre-existing privacy laws. Accordingly, there legal, technological, and practical implications of applying two different regulatory schemes to the same data flows presents many issues worth considering.

This outline focuses on the provisions in the pre-2023 privacy laws that limit “profiling” or “automated decisionmaking” (collectively, “ADM”). This Brainstorming Group chose these provisions because they have a natural technological nexus with a wide variety of AI systems. Additionally, privacy enforcement strategies by government regulators (particularly the Federal Trade Commission) have included algorithmic disgorgement. If applied against the AI systems that have become available since 2023, this method of enforcing the privacy laws could have far reaching impacts on AI adoption strategies. In short, privacy compliance right in instances ADM is a requirement for an organization that wants to use AI systems to those ends.

In many of our meetings we wrestled with the implications of new AIs and their use cases. These far reaching discussions led us to the conclusion that if we were to complete an outline, it had to focus on a particular set of facts. Accordingly, we built a fact pattern based on the December 2023 enforcement action by the Federal Trade Commission against *Rite Aid*. The case involved algorithmic disgorgement, facial recognition technology, and an automated system that prompted a range of actions for workers in Rite Aid stories (i.e., limited human

decision-making). The resulting facts spurred a robust discussion about the interaction between the ADM provision in privacy laws and the growing body of AI law. Hopefully the richness of that discussion comes through in the outline.

This discussion, though, led to many more questions about the interaction of the pre-2023 privacy laws and the emerging AI laws. The last section of the outline tries to organize these ideas around concepts in the pre-2023 privacy laws that may lead to challenging results in the context of AI systems and emerging AI laws. There are undoubtedly countless other topics to explore, but these are the ones that arose from the discussion about how ADM provision in the privacy laws interact with AI systems and emerging AI laws.

If this effort leads to a drafting team, we believe it could be one section in a Sedona commentary on the intersection and interaction of privacy law, AI systems, and AI law.

Fact Pattern

A retail company with multiple brick-and-mortar stores employs an automated system to help identify people coming into the store who present a security risk to the business. The automated system does this by applying facial recognition system to scan CCTV footage. If there is a potential match, the system generates alerts to in-store employees on company-provided devices, like an app on mobile phone.

The facial recognition system works by comparing CCTV footage of all customers to enrolled images. Enrollment happens in a few ways. Primarily, store employees who witness poor customer behavior capture a photograph of the individual. This capture happens by extracting images from CCTV footage, taking photos with company provided devices (e.g., mobile phone), and uploading photos from police BOLO alerts or news reports about local crime. With all these methods, an individual does not necessarily know her image is being captured for enrollment purposes. On occasions where an individual does know her image is being captured, she is also more likely to provide other information relevant to the enrollment – her name. But employees also capture the person's name by photographing the person's identification card (i.e., driver's license, work badge, etc.), without letting the person know about the enrollment.

Other information included in the enrollment includes the location of the store where the enrollment is taking place and the reason for the enrollment (e.g., suspected of shoplifting, belligerent behavior towards staff, suspicion of violence, etc.). An enrollment is completed once the facial recognition system generates a "faceprint." Enrollment in the system can be completed even if the employee does not provide the individual's name.

These alerts include notice of the potential match, a copy of the enrollment photo, a still image captured from the current CCTV footage, a confidence score about the match, the name of the purported individual (if it is available), and instructions to the employee on how to proceed with the individual. Once the employee completes the assigned task, they can log the alert as "resolved" and provide notes. There are four possible instructions to the employee.

1) Observe from a Distance – The employee should be aware of where the individual is throughout their visit to the store. The person should be treated as suspicious, but not approached unless they are engaging in suspicious or threatening behavior. This alert is

resolved by either confirming that the individual was observed throughout their visit or by noting that the individual engaged in suspicious or threatening behavior.

2) Observe and Provide Customer Service – The employee should approach the individual within one minute of the person entering the store. If the person is not engaging in suspicious or threatening behavior, the employee should offer assistance to the individual. Then the employee should observe the individual throughout their visit to the store. This alert is resolved by either confirming that the individual was observed throughout their visit or by noting that the individual engaged in suspicious or threatening behavior.

3) Ask to Leave the Store – The employee should approach the individual as soon as safely possible and ask the individual to leave the store. The employee may ask the person to identify herself, but this is not necessary. If the individual asks why, you may say that the person has been suspected of behavior detrimental to store or its employees. If the individual does not leave after being asked, the employee may call law enforcement, at their discretion. The employee, however, should not reveal to any other party that the facial recognition system generated the alert. The alert is resolved after the employee has attempted to ask the individual to leave the store. It should be noted if the person left the store or if law enforcement was contacted.

4) Notify Law Enforcement – The employee should not approach the individual. The employee should notify law enforcement promptly that an individual needs to be removed from the store. The employee, however, should not reveal to any other party that the facial recognition system generated the alert. The alert is resolved after the employee has contacted law enforcement. It should be noted if law enforcement asked for a specific reason to remove the individual.

The resolutions from the alert system are not incorporated into the future operations of the facial recognition system. An employee's rate of resolution or quality of resolution are not reviewed or evaluated. As a result, there is not a mechanism whereby store employees identify false positives. The system is not set up to improve from later human inputs. Accordingly, whatever bias and issues that exist in the facial recognition system will continue to persist.

Topics to Explore Related to Automated Decision-making/ Profiling

Presence of PII

The presence of PII makes the privacy laws apply. Assuming that the AI system is processing PII, how the PII is used is the subject of AI regulations.

Other questions

How are the anonymization/pseudonymization standards impacted by the use of PII in AI systems? Does it present an increased risk of reidentification, thereby removing the "privacy law safe harbor" offered by anonymization/pseudonymization provisions?

Even if an AI system does not specifically process PII, is there still a risk that it will generate information about "an identified or identifiable person" by inference? Are there technical/administrative safeguards that can reduce the chance of the AI system alone, or in

combination with other data from the controller (or others), generating PII? If so, to what extent should the privacy laws apply to such an AI system and its outputs?

Automatic Decision-making / Profiling

The privacy law provisions about automatic decision-making or profiling (collectively, "ADM provisions") present as technology neutral. Rather, they focus on the goals of processing the PII and the organizational behaviors that result from such processing. This means many different use cases of AI within organizations could qualify under the ADM provisions.

EU Law

Article 22(1) of GDPR prohibits "a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." The facial recognition system in the fact pattern qualifies as automated processing because the result could cause legal effects like removal from the store and limiting future access to the stores.

Under the EU AI Act Art. 7, "high risk" AI systems are those that "pose a risk of harm to the health and safety, or a risk of adverse impact on fundamental rights," or fall into any of the following categories (Annex III): biometrics; critical infrastructure; education and vocational training; employment, workers management and access to self-employment; access to and enjoyment of essential private services and essential public services and benefits; law enforcement; migration, asylum and border control management; and administration of justice and democratic processes.

Many scenarios that qualify under the ADM provisions would also qualify as "high risk" under the EU AI Act because of the commonality of impact on fundamental rights. The story in the fact pattern provides one example. Even if sufficient human involvement means that the ADM provisions do not apply, the AI system itself may still be considered "high risk" under the EU AI Act.

The Brainstorming Group believes that it is safe to assume that any instance that would qualify as automated decision making under GDPR Art. 22(1) would meet the definition of a "high risk" AI system under the EU AI Act.

From a compliance perspective, what would full compliance look like for "high risk" AI system under both the EU AI Act and GDPR? What are the types of administrative/technical controls that allow such data processing to occur lawfully under both laws? Is there a meaningful difference in complying with the EU AI Act only, and not GDPR, and vice versa?

U.S. Laws

At least California Consumer Privacy Act (CCPA) and the Colorado Privacy Act have ADM provisions that are in effect. Under both CCPA (1798.185(a)(16)) and the Colorado Privacy Act (6-1-1306)(1)(a)) a consumer has a right to opt out of "profiling." The facial recognition system fact pattern meets the respective definitions of "profiling" (CCPA 1798.140(z) and Colorado 6-1-1303(20)) because this system uses "automated processing to .. analyze … [an individual's] … location."

The emerging U.S. laws on artificial intelligence seem to be more narrowly tailored than the EU AI Act. Accordingly, there will be comparatively fewer occasions for a business operating in the

U.S. to have to comply with both a privacy law that includes ADM provisions and an AI law governing the same data processing. For example, New York City enacted the Automated Employment Decision Tools (AEDT) which targets certain AI systems used in the hiring context. But, there is no corresponding comprehensive data privacy law with an ADM provision in either New York city or New York state.

One state where the privacy law and AI specific-laws will interact is California. The California Consumer Protection Agency is drafting CCPA regulations about ADM. These regulations incorporate AI systems into the definition of automated decision-making under CCPA. Accordingly, the full complement of requirements under CCPA (and possibly additional ones) would apply to PII being processed by AI systems.

From a compliance perspective, what are the differences between the administrative/technical controls that allow such data processing to occur lawfully under CCPA for normal data processing versus CCPA when an AI system is involved? Considering the rapid rate of AI growth, will this bias entities towards compliance with the "AI specific" regulations under CCPA, out of an abundance of caution?

Privacy Law – Consent/Opt-out

One of the lawful bases for ADM under GDPR (Art. 22, Recital 71) is "based on the data subject's explicit consent." Explicit consent can be withdrawn at any time. The California and Colorado privacy laws allow ADM to occur until a data subject "opts out." Under both bodies of law, accordingly, consent is temporary.

If an AI system is being used for ADM, what does withdrawal of consent at the following moments look like from a technological and administrative perspective?

- PII enters the AI system (effectively preventing PII from entering the system)
- the AI system generates a decision/profile
- Whether/how a human being uses that AI-generated decision/profile? (reinforcement learning models)
- After a decision has been made about a data subject

This leads to a key question about the applicability of privacy laws to AI systems - when does the ADM happen? Does it happen only when a human takes action based on the system's decision? Does it happen when the AI system generates a decision? Does it happen when PII enters the system, and therefore is eligible for these later actions? Technological specifics may be a factor in making this determination. But, from a compliance perspective, a controller may need to decide for itself when ADM happens, and develop appropriate safeguards accordingly.

Assume a data subject opts out of ADM. At a minimum, it seems that the AI system could not be used to make decisions about that data subject? From a privacy law compliance perspective, would the organization need to maintain decision-making/profiling process that completely avoids the AI system, or simply injects enough human-considered factors to avoid the statutory definition of ADM? Does this bias organizations against using comprehensive AI systems, like

Microsoft's Copilot? In practical terms, would the human "workflow" for such a process match the disclosure requirements under the AI laws of how the AI system processes operates?

GDPR – Other lawful bases for ADM

Considering the challenges presented by the withdrawal of consent for ADM in the context of an AI system, organizations subject to GDPR may seek other lawful bases for ADM. Under GDPR Art. 22, these other lawful bases are if ADM:

> "is necessary for entering into, or performance of, a contract between the data subject and a data controller; or
> is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests."

From an administrative and technical perspective, how can organizational discipline be maintained to ensure that only PII subject to these lawful bases is used in the AI system for ADM purposes?

Both contracts and other legal authorizations which support ADM have implicit time limitations (e.g., contract period of performance, statute of limitations, fulfilling a specific statutory purpose). Accordingly, if an AI is being used for ADM under these lawful bases, what happens to the PII when that time limit is reached? Though that specific data subject may not be subject to ADM, does the AI system need to generate different outcomes as part of demonstrating compliance with the time limitations inherent in these legal bases for ADM? If not, what other features of the AI system need to demonstrate safeguards of "the data subject's rights and freedoms and legitimate interests"?

Please note that under the U.S. privacy laws these other lawful bases do not seem to be an option. A consumer's right to withdraw consent for ADM seems to be absolute. But, some sectoral specific laws (e.g., insurance regulations, health data processing rules) may create instances which give rise to a similar analysis.

Other Topics to Explore at the Intersection of Privacy Laws and AI Laws

Many data privacy laws embody the fair information privacy principles The emerging AI laws seem to apply similar principles. In our discussion, the Brainstorming Group identified key areas of conceptual overlap that may merit further exploration.

Right to Delete Data
What happens if a data subject exercises her right to data deletion under the privacy laws and requests that her data be deleted after it has been placed in an AI system?

It is possible to interpret the definition of PII in the context of an AI system to incorporate not only the data directly attributable to the data subject, but also the inferences/learning the AI system has gleaned from that data subject's PII. In principle, therefore, AI systems may need to

be "versioned" to each moment a new data subject's PII enters the system. That way, withdrawal of consent does not run afoul of the privacy laws.

This may be impractical, if not technologically impossible. Accordingly, what are the practical steps that a controller must take to demonstrate the removal of a data subject's PII from an AI system?

Transparency
Transparency in the privacy law context generally mandates that a controller describe what PII will be collected and for what purposes. In practice, particularly in the United States, privacy notices have become very generic, almost to the point of being rote.

The AI laws seem to mandate transparency as well, particularly as to explaining how the AI systems work. Under the privacy laws, sharing of this kind of information would generally not be required, as it might be considered a trade secret information and therefore not PII. But, some AI systems, or elements of them, are simply not explainable.

Several questions come to mind. Do the transparency requirements of the AI laws go beyond those of those of the privacy laws? If so, can an organization that complies with the AI's transparency requirements be understood to be "oversharing" under the privacy laws, and therefore reduce their compliance burden? Conversely, does failure to explain an AI system under the AI law also run afoul of the transparency requirements in the privacy laws?

Purpose Specification
In many respects, compliance with the privacy laws hinges on an organization's commitment to honoring the purpose it specifies for collecting and processing PII. An organization, accordingly, could indicate that a data subject's PII might be used to train an AI system. Aside from training, assume the AI system may never process that data subject's PII.

If the controller then "sells" the trained model, or offers access to third parties, does that constitute a different purpose for the PII under the privacy laws? Relatedly, would that constitute a data sale under the privacy laws (*e.g.*, the Sephora decision under CCPA)?

Some AI systems are self-learning. Assume an AI system starts operating beyond a point at which the organization can explain, as required under the AI laws. The four principles of explainable AI proffered by the OECD would suggest that AI systems should be able to be "stopped" once they operate beyond their point of explainability. Would such an event constitute a violation of the principle of purpose specification under the privacy laws?

A common problem sits at the heart of both these scenarios – under the privacy laws, is the PII in the AI system and the processing done by the AI system one and the same? Or, are they distinct?

U.S. FTC enforcement actions of privacy harms resulting from PII being fed into algorithms have involved algorithmic disgorgement (see this fact pattern based on the *Rite Aid* case). This would suggest that under the privacy laws, the PII and the AI system are not distinct. The *Uber* cases in EU (here and here) suggest that as long as organizations allow for adequate human

involvement in decision-making, AI systems can process PII extensively with little fear of running afoul of the ADM provisions in GDPR.

Are there technical and administrative safeguards and practices that an organization can employ to avoid having to ensure compliance across both regimes?

Risk Assessment in the Context of Breaches

In the data breach context, part of risk assessment involves articulating known risk. The emerging AI laws also require articulation of how an AI system works and the known risks. This conceptual overlap gives risk to some questions.

For those data privacy laws that incorporate risk assessment into the breach notice standards, would an unauthorized disclosure to the AI system described in the fact pattern (*i.e.*, accidental enrollment) present an unacceptably high risk? Imagine that an applicable AI law has an explainability requirement. Can the "explainability" of an AI system serve as a guideline for when unauthorized disclosure to an AI system presents an unreasonable risk in the data breach context?