Security & Privacy in AI

Input into 27091: Security and privacy in ai, ad-hoc committee

Submitted by Mark Lizar
ANCR WG Co-Chair & Editor Submission
* To be distributed to WG 5 Mirror Committee 28 February 2024

References (in progress)

Normative Law

Council of Europe - Convention 108 + (Conv.108+).

- Conv.108+ Art 14 (1-8) mirrored by GDPR, Art 12 (1-8)
- Conv.108+ Art 31 Records of Processing Activities (RoPA) Mirrored by GDPR Art 30 Records of Processing Activities (RoPA)
- Conv.108+ Logging Art 88.

Non-Normative Law

- Bill 64 (2021, chapter 25), An Act to modernise legislative provisions as regards the protection of personal information, came into force September 23, 2023. Now referred to as Law 25, An Act to modernise legislative provisions as regards the protection of personal information, SQ 2021, c 25.
    - Canada, Quebec Law 25, in which Secondary purpose for consent is legislated, although not specified explicitly in GDPR or Conv. 108+. The principles of data portability and personal data control cover the explicit rendition used here.

International Standards

- 29100 – Security and privacy framework (open access) interoperable with international regulation and international organization security standards ISO/IEC 27001
- 29184 – Online privacy notice and consent: Including (2020) – Consent notice receipt in Annex B
- 27560 – Consent record information structure (2023) an adaptation of the Kantara Consent Receipt V1.1.

Internet Standard

- W3C Data Privacy Vocabulary.

Non-Normative

- Kantara ANCR WG
    - Digital Transparency Performance Scheme, Part 1 and 2[1] v0.9.1 (in comment review)
        - Presents both a conformance assessment and compliance assessment methodology
        - Specified with above normative references
    - ANCR- Auth C- Protocol for Digital Privacy

[1] Digital Transparency Performance Scheme, ANCR WG, Kantara Initiative, Sept 15, 2023 [Online: https://kantara.atlassian.net/wiki/spaces/WA/pages/301564731/ANCR+Transparency+Trust+Performance+Assessment+Scheme]

        - Notice, notification, disclosure etiquette for using notice and consent, consent receipt-based governance
        - Technical consensus exchange protocol for authorization from consent (AuthC)
    - 0PN-ZKP: Individual remains anonymous and uses validated claims (a cyber-notary) in which the Controller is identified and sousveilled, enabling:
        - Authentic consent tokens
        - Core flow of 0PN-ZPN is explained below.

ANCR Digital Transparency & Trust Performance Scheme

- ANCR scheme is used to capture the transparency state in a standard record format specified to ISO/IEC 29100 privacy framework standard.
- The record can subsequently be assessed and audited for transparency adequacy, security conformance, and privacy compliance
- For ease of writing, Convention 108 + is used as the authoritative legal document for referencing what is demonstrated as an Adequate Digital Transparency Modality, or a Compliant Digital Transparency Modality to demonstrate Adequacy for transborder consent with reference to
    - Conv. 108+ Art 14 and GDPR Art 12.

Non-Normative Terms and Definitions

[Editorial Note] These terms and definitions are in addition to the existing normative framework, and in draft mode for comment and review.

Authority

- Referential to the authority in which the stakeholder presents a claim or requirement for interaction in accordance with the transparency state

Claim

- Referential to a proof of notice receipt, which is in effect a legal claim, that Is used instead of an 'I Agree' Button – contract agreement framework, but instead utilizes a privacy agreement framework, utilizing notice and consent (a human process) rather than the technical process for interoperability and exchange.

Consent Notice Receipt (or Consent Notice Receipt: ISO/IEC 29184 Appendix B)

- A linked consent record providing the PII Controller Identity, Contact, and Summary of Purpose and Associated Categories for processing provisioned by the controller or processor.

Consent Token

- A consent receipt which is used as a verified claim token is sent back to point of origin to exercise rights, regardless of the legal justification for surveillance.

Credential

- A document, record, or receipt, that presents an attribute, or multiple attributes for use as an identifier.

Digital Credential

- Claims in a consent receipt that are signed with cryptographic keys for secure exchange of a record or receipt between parties

Micro-Receipt Credential

- A credential generated from a PII Notice Controller or Provider, in whatever context, physical or digital, the notice is being provided in.

PII Controller Notice Credential

- The minimum set of required data attributes to digitally identify 1. The authority of the PII Controller; the identity of the Controlling entity; the privacy contact point for access to privacy services; the contact details of the data privacy officer for direct inquiries or complaints; the address, legal jurisdiction, regulators contact, complaint, and their (legal/governance rules) for engaging with privacy rights-based governance
- Identity of privacy elements that replace the physical, including:
    - To use the notice / transparency in a privacy operational manner
    - To meet required minimum - PII Controller Notice Attribute requirements to conform to security and privacy requirements and comply to a human expectation of digital privacy.

Secondary Purpose for Consent

- Any service that collects, processes, surveils, infers, or captures personal data for any legal reason provides consent receipts to enable not only consent controls but also, to enable secondar consent, in which the PII Principle, directly consents for the use of the data for a secondary legitimate purpose

Introduction

The Kantara Initiative Digital Trust, Transparency Performance Assessment Scheme

This scheme is used to benchmark the performance of AI transparency in real-time, for consent and data control. The aim is to develop this scheme into an international digital consent conformance and compliance scheme for data transfer.

The Use Case Summary

In simple terms, whenever AI is being used, a notice is required in which permission is requested, the AI algorithm is identified and receipt provided with the Controller, as a legal requirement.[2] This two-factor

[2] Identity of the controller, GDPR Article 13.1 and in Council of Europe 108+ Art 15.1

concentric notice is used to produce a record of the Controller and the algorithm being used, which is provided to the individual as a consent token receipt.

A consent receipt is used to make consent tokens that are sent back to the AI service (the digital endpoint) to enable individual consent and data consent controls. For example, to withdraw consent for the algorithm to use that personal information in its Large Language Model (LLM), or to provide a third-party consent token to use the Personally Identifiable Information (PII) along with the generative AI LLM. In accordance with data portability and control requirements.

Or conversely, in proportion to the use of the generative AI technology. Distributing data governance to the individual. As the record of the notice is legal evidence of the consent policy, which can be assured by the Notarial consent protocol AuthC, to provide levels of assurance higher than currently provided by personal identifier-based security and data surveillance systems.

Importantly, Canada has a new law that has recently come into force, which explicitly states that the individual can consent to a secondary purpose, regardless of the legal justification of the surveillance. The consent receipt can then be used as a secondary purpose consent token, to facilitate digital privacy right controls, and by so doing govern the surveillance of AI, age assurance, digital twins, IoT, and Video Surveillance[3], with the same normalized of technical practices which can now be co-regulated using a common record information structure. As published in ISO/IEC 27560:2023 Consent record information structure.

Decentralizing Data Governance

The individual can use this receipt to see how often, or even when, their content and information is used in a generative AI, enabling decentralised generative AI governance.

As is identified there are significant ethical issues with ungoverned AI creating tremendous security risks, addressed with this ISO/IEC specified notarial record of processing activities.

1. LLMs are being created with scraped data without consent
2. Subsequently, they are used to make products, such as synthetic data models, without consent
    a. This enables the modelling of individual behaviour for research and development
    b. All of which is personal, private, intimate, and violates the autonomy of the individual
3. Since the Digital Services Act came into force (February 17 2024), these mechanisms can be used in the market to see the scraping of personal data (using cookies and pixels) as illegal and a violation of privacy and an individual's security
4. Secondary Consent – a consent receipt provides for a secondary consent purpose. Implementing this requirement forces the decentralization of data governance as it enables people to take control of the data that is out there already.

[3] Surveillance Trust is an implementer for physical access points https://www.surveillancetrust.org/

Trust Performance Benchmarking Scheme for Digital Transparency

ANCR Transparency Performance Scheme (TPS) is an assessment that indicates to the individual if personal data is being scraped before consent, e.g. before the identity for the Controller is provided. If the identity was complete, if the identity is accessible according to context, and if the digital security or trust risks are presented; especially the transborder transfer of personal data identifiers.

In Kantara we have a model TPS, with 2 parts. Within Part 1 there four transparency Indicators which are used to create a controller record and assess its digital trust compliance. In the appendix part 2 of the scheme is an example of how it can be applied. The aim is to develop this into an ISO/IEC Benchmarking Scheme, for an international digital consent trust assurance.

The scheme provides a standard way to measure the compliance and conformance, decentralizing and distributing the governance of all identifier surveillance technologies in this regard applicable to the security and privacy of AI, age assurance and digital twin technologies.

Use Cases

The use cases presented are in the context of assessing the 'Adequacy' of security and privacy digital transparency, by generating an anchored record of the processing activity, called a notice receipt. This creates a record of the processing activity using the ANCR Transparency performance scheme for digital identity trust.

Two use cases are presented in which a proof of knowledge (consent receipt) is generated; either when a notice is read and an action to accept this notice is taken or by the service through a presented PII Controller Identity.

In an operational privacy scenario (not covered by these use cases) A recorded fingerprint of the digital security and privacy state. Is kept by all stakeholders.

1. Assess the conformance of the controller identity record structure and format according to ISO/IEC 29100 and using the Kantara Consent Receipt in ISO/IEC 29184 Online Privacy Notice

Use Case Overview

A record of a processing activity is created in part 1 of the scheme, by applying the four Transparency Performance Indicators, as specified. (Scheme 1 Part 1)

- Create a standard record of the PII Controller, and the state of security and privacy transparency.

Scheme 1, Part 2

- To use the conformant record to assess the compliance of the PII Controller transparency, which can be done in many ways.

Use case 1 description:

This party is Data Protection Officer, Regulating Authority, auditing or actively monitoring the security and privacy of AI - real-time the state of digital privacy.

In this use case the TPI Scheme 1 is applied as instructed above to set up a benchmark program for security and privacy in AI use cases.

Use case 2 description:

The auditor is a PII Principle (or master data controller in a Consent Token wallet) in which the consent button, and access to all the data sources where secondary consent functions and features, sit. In this use case, the individual can provision their own consent receipt, which is called a consent token, and use the TPS scheme Part 2. to measure the performance of the service conformance with ISO/IEC 29100 and compliance with GDPR and Conv. 108+, autonomously.

Levels of Authentication Trust Assurance

Additional levels of authentication assurance can be added which match the current levels of authentication assurance provided by NIST 800-63, through numerous methods, and in so doing surpass current security assurance, while removing the need to directly surveil the identifiers of a PII Principal.

The PII Principal (master controller in this instance) is the creator of the consent token, enabling the human, not the service, to manage their own consent that can further be assured by authorised third parties, such as data privacy officers for authorities, or data privacy regulators.

The resulting assurance framework enables digital consent by providing the individual the technical capacity to issue their own consent tokens, a non-fungible-token (nft), to decentralise governance of surveillance-based technologies.

Use case 3: (not in scope)

Mirrored Records- In practice – the security and privacy state are captured in use case 1, and is generated by wallet in use case 2, to mirror the coinciding record that is created by the PII Controller, when a credential or its registry API is used. This digitally twinned record is then considered anchored by the individual in the receipt wallet when the consent receipt token is issued, thus enabling the individual to see the data processing that occurs when the token is used.