Submission AI Use Case from

Mark Lizar – Co-Chair and Editor of ANCR WG- Kantara Initiative

This submission, towards a NWIP proposal for the outlined ANCR Transparency Conformance and Compliance scheme. This NWIP would be focused on security, privacy and ethics use cases, concerning the governance of AI technologies, with a privacy and security architecture consisting of A) personal AI, for personal data control B) a Public Commons AI, consented/legitimate data surveillance C) commercial/institutional AI in which the LLM is required to provide digital privacy rights for legitimate data processing, and enable secondary purpose with directed consent from the PII individual. It is proposed that governed AI will greatly expanding the utility of Gen AI services for people, (children, youth and communities) increase genAI trust through transparency over PII processing.

# User story

In the ANCR Governance framework there are 3 Vectors of data control based governance that make up the architecture

1. Personal Data Control
2. Data Protection (Commercial and Institutional
3. Co-Regulated – utilising the international Digital Privacy Transparency

As a PII Principle, I want to ensure that I have a receipt, generated independently of the service so I have clean data record of the digital identity relationship with an AI services, so that I can manged and control my PII used and accessed by a genAI service. The PII Principal has a secure consent notice receipt repository for a secure private AI,

## Template

|  |  |
| --- | --- |
| Category | Gen AI – Violating Privacy – A) LLM Created with Secondary Scraped Data Non-Compliance – Unethical AIB)– taking the inputs – adding to LLM without Consent when using AI - Threat to Privacy  |
| Topic | Compliance Consent, Ethical AI, Non-Biased AI, Accountable AI, (private, public and common AI)  |
| Story | As a [stakeholder], I want to be assisted on [securing and protecting my PII] so that [I can choose what AI and what services benefit my data] |

## Examples

|  |  |
| --- | --- |
| Category | Security and Privacy of Gen AI  |
| Topic | Standard development |
| Story | As a [assessor of gen AI transparency compliance ] I want to be able to see the providence of the information in an AI, to understand the level of bias, security and privacy risk, as well as the liability of using an AI service. How does a stakeholder know if a LLM’s (Large language Models) is a data model that is created through mass aggregation of human contributed data, including personal data, which is the use of personal data for a secondary purpose, without consent. Services using personal data for any secondary purpose, including differential privacy, is a secondary purpose and requires consent,  |

|  |  |
| --- | --- |
| Category | genAI data-breach |
| Topic | Data governance and compliance |
| Story | As a privacy stakeholder in a GenAI service. I want transparency over the Controller Identity, who benefits from the GenAI Service, What Governance Framework (laws) the GenAI service adhere’s too.  |

|  |  |
| --- | --- |
| Category | GenAI as a threat to privacy |
| Topic | Digital Privacy Transparency  |
| Story | As a user of a GenAI service, I want to receive a consent notice receipt when engaging with the service so that I can control the PII Collected during interaction, and I want every GenAI query to be logged and linked to this GenAi, to receive a record of processing activity. The record of processing activity can then be use by my own private AI, which only works with these records, and no other sources. Enabling digital privacy transparency. This transparency can be realized by using a receipt to query the AI to see what PII the GenAI service has, and to withdraw consent, object to it’s processing or delete the PII data. As well as to query or be notified when that PII has be access by a Gen-AI \*\*\* As a [GenAI provider], I want to be scrape as many data sets and sources as frequently as possible into the LLM, on [the analysis of a dataset], without a log or any records of processing activities, so that I have access to query [PII on a specific person so it can be derived] without a log file or linked record, so the PII Principle or regulator has no transparency over when their data is used or collected. Then when the AI is used by a PII Principle, the personal micro-data (also referred to as meta-data) can be collected and correlated real-time. |

## Example of topics

|  |  |  |
| --- | --- | --- |
| Domain | Topic | Source |
| IT workforce | * Text translation and productivity
* Democratize organization data
* Easy to generate SQL queries
* Chatbot for dialog and prompt engineering
* Code migration
* Easing management meeting
* Code quality improvement
* Software coding
* Generating images with AI
* Computer programmer and data scientist
* Security audit will be easy
* Mock interview
* Apps creation with generative AI mode
* Time management
* Benefits for editing
* Behaviour prediction and creative writing
* Fact checking, generating ideas
 | Unleashing the Potential: Overcoming Hurdles and Embracing Generative AI in IT Workplaces: Advantages, Guidelines, and Policies. Pan Singh Dhoni |
| Helping cybersecurity | * Malware detection
* Anomaly detection
* Password cracking
* Threat intelligence
* Adversarial AI defense
* Phishing detection
* Network traffic analysis
* Automated security response
* Security training and simulation
 | Synergizing Generative AI and Cybersecurity: Roles of Generative AI Entities, Companies, Agencies, and Government in Enhancing Cybersecurity. Pan Singh Dhoni  |
| GenAI against cybersecurity  | * Fund raising
* Ransom email (layman can do it)
* Generate malicious domains
* Phishing kits
* Circulating manipulated photos
* Fake digital content
* Fake voice
* Fake document generation
* Compromising company intellectual property
 |
| Enhance data quality | * Data imputation and completion (Completing missing data, Time series forecasting)
* Data validation and cleansing (Data anomaly detection, data quality validation, data standardization)
* Data augmentation (Enhancing datasets, Feature engineering – privacy?)
* Data simulation and testing (Realistic test case, stress testing)
* Data governance and compliance (Data masking, auditing and monitoring)
 | Exploring the Synergy between Generative AI, Data and Analytics in the Modern Age. Pan Singh Dhoni  |

# Use case

## Template

|  |  |
| --- | --- |
| Use case name |  |
| Ecosystem | *Describes the ecosystem: identifies the systems of interest, the stakeholders, and the stakeholders’ assets that are impacted by GenAI* |  |
| System of interest: < Use case system of interest >  |
| Assessment of system of interest | *Assessment on security and privacy concerns* |  |
| Security and privacy concerns | *Highlights security and privacy concerns that are impacted by GenAI* |  |
| Security and privacy risks | *Identifies security and privacy risks that are impacted by GenAI* |  |
| Security and privacy controls | *Identifies security and privacy controls that are impacted by GenAI* |  |
| Security and privacy assurance | *Identifies security and privacy assurance aspects that are impacted by GenAI* |  |
| Security and privacy plan | *Identifies security and privacy plan aspects that are impacted by GenAI* |  |

## Example (from Yunwei Zhao -CN)

|  |  |
| --- | --- |
| Use case name | Generative artificial intelligence manual annotation and training |
| Ecosystem | *Describes the ecosystem: identifies the systems of interest, the stakeholders, and the stakeholders’ assets that are impacted by GenAI* | * Systems of interest:
* Generative artificial intelligence system
* Data annotation system
* Stakeholders:
* Generative artificial intelligence service provider
* Annotation service provider
* Third-party evaluation agency
* Generative artificial intelligence user
* Stakeholder assets that are impacted by AI
* Annotation tools
* Annotation data set
* Generative artificial intelligence training data set
 |
| System of interest: Generative artificial intelligence system |
| Assessment of system of interest | *Assessment on security and privacy concerns* | * Security concerns on generative artificial intelligence system are significant.
* Privacy concerns on generative artificial intelligence system are significant.
 |
| Security and privacy concerns | *Highlights security and privacy concerns that are impacted by GenAI* | * All security and privacy protection goals to consider for generative artificial intelligence system (confidentiality, integrity, availability, unlinkability, transparency, intervenability)
* All security framework concepts to consider for generative artificial intelligence system (Identify, Protect, Detect, Respond, Recover)
* All privacy framework concepts to consider for generative artificial intelligence system (Identify-P, Govern-P, Control-P, Communicate-P, Protect-P)
 |
| Security and privacy risks | *Identifies security and privacy risks that are impacted by GenAI* | * Privacy risks related to generative artificial intelligence system (e.g., disclosure of identity information and sensitive information etc. while performing generative artificial intelligence training and content generation)
* Security risks related to generative artificial intelligence system (e.g., The annotated data used for model training is incorrect, inconsistent, or subjectively biased, resulting in harmful, deceptive, or biased generated content)
 |
| Security and privacy controls | *Identifies security and privacy controls that are impacted by GenAI* | * Security controls from ISO/IEC 27002 should be considered for generative artificial intelligence system (e.g., information security policies, asset management, physical and environmental security, access control, operation security, information security incident management)
* Privacy controls from ISO/IEC 27701 should be considered for generative artificial intelligence system
 |
| Security and privacy assurance | *Identifies security and privacy assurance aspects that are impacted by GenAI* | * Organization using generative artificial intelligence system to ensure that system can be audited (sees ISO/IEC 26000). This includes organizational and technical evidence.
 |
| Security and privacy plan | *Identifies security and privacy plan aspects that are impacted by GenAI* | * Organization using generative artificial intelligence system to establish a security plan, that will be validated and reviewed periodically for continual improvement.
 |
| System of interest: Data annotation system |
| Assessment of system of interest | *Assessment on security and privacy concerns* | * Security concerns on data annotation system are significant.
* Privacy concerns on data annotation system are significant.
 |
| Security and privacy concerns | *Highlights security and privacy concerns that are impacted by GenAI* | * All security and privacy protection goals to consider for data annotation system (confidentiality, integrity, availability, unlinkability, transparency, intervenability)
* All security framework concepts to consider for data annotation system (Identify, Protect, Detect, Respond, Recover)
* All privacy framework concepts to consider for data annotation system (Identify-P, Govern-P, Control-P, Communicate-P, Protect-P)
 |
| Security and privacy risks | *Identifies security and privacy risks that are impacted by GenAI* | * Privacy risks related to data annotation system (e.g., disclosure of identity information and sensitive information etc. while annotators performing annotation tasks or caused by insecure annotation tools)
* Security risks related to data annotation system (e.g., annotators maliciously annotate data, unsafe annotation strategies lead to increased data risk during the training phase, resulting in greater security and ethical issues)
 |
| Security and privacy controls | *Identifies security and privacy controls that are impacted by GenAI* | * Security controls from ISO/IEC 27002 should be considered for data annotation system (e.g., information security policies, asset management, physical and environmental security, access control, operation security, information security incident management)
* Privacy controls from ISO/IEC 27701 should be considered for data annotation system
 |
| Security and privacy assurance | *Identifies security and privacy assurance aspects that are impacted by GenAI* | * Organization using data annotation system to ensure that system can be audited (sees ISO/IEC 26000). This includes organizational and technical evidence.
 |
| Security and privacy plan | *Identifies security and privacy plan aspects that are impacted by GenAI* | * Organization using data annotation system to establish a security plan, that will be validated and reviewed periodically for continual improvement.
 |