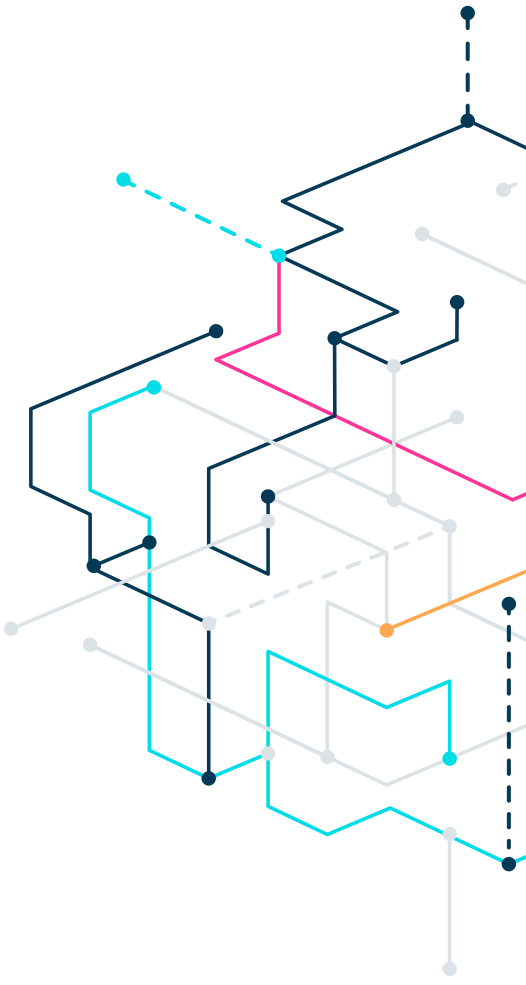
anecdotes

The anecdotes AI Toolkit

A comprehensive list of risks, controls and policies to securely integrate Generative AI into your organization.

Index



Author

Ethan Altmann,
Compliance Product Owner @ anecdotes

Contributors

The anecdotes AI Toolkit was created in collaboration with this exceptional group of industry leading experts who have each lent their unique perspective and experience to this document. Their invaluable insight has enabled the vision of a practical toolkit for Security and GRC professionals to come to life.

Val Dobrushkin
Director, Risk and Compliance @ NoName

Prabhath Karanth
Global Head of Security & Trust @ Navan

Karl Mattson
CISO @ NoName

Esther Pinto
CISO @ anecdotes

Tamir Ronen
Global CISO @ HiBob

Olivia Rose
CISO @ Rose CISO Group

Dineshwar Sahni
Senior Cybersecurity Leader

Ronit Shlyfer
Information Security & Compliance, Team Lead @ Riskified

Omer Singer
Head of Cybersecurity Strategy @ Snowflake

Matt Szymanski
Director of Security Engineering @ Yext

Tyler Young
CISO @ BigID

Preamble

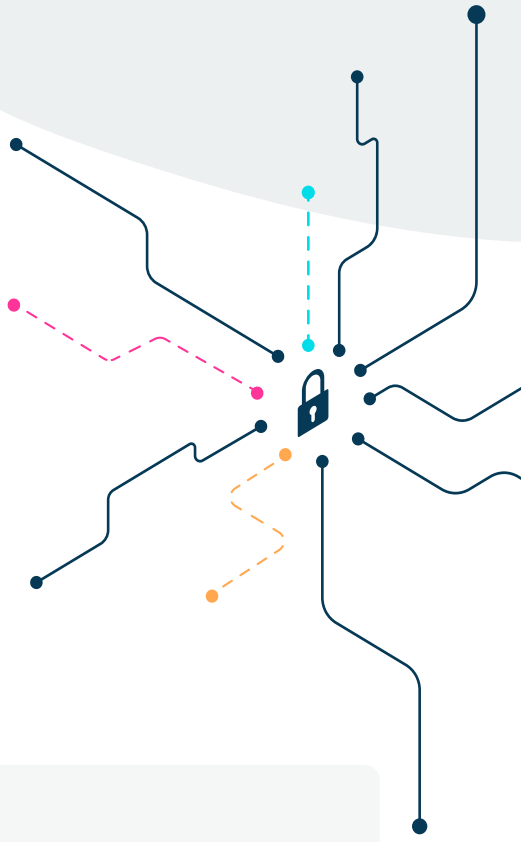
New GenAI use cases are being discovered, tested and brought into the limelight on a daily basis. Striking the right balance between justified excitement and a healthy dose of skepticism is a challenge faced by countless tech leaders, but especially by Security and Compliance professionals. The core challenge they face lies in empowering their organization to reap the benefits of this new technology, whilst ensuring that the new processes do not exceed the organization's defined risk appetite.

This document aims to give organizations a clear framework through which Generative AI can be securely integrated into existing organizational operations, without impeding security and Compliance processes and obligations. By implementing the framework, security and Compliance leaders take an important step towards ensuring that their organization remains at the forefront of embracing modern technology, whilst refraining from excessive risk exposure and conforming to best practices.

Practically speaking, this toolkit takes a framework based methodology by which organizations can securely and effectively use and implement GenAI tools:

- By understanding the associated risks and their potential-impact on the organization
- By understanding and implementing appropriate mitigative controls to reduce these risks to an acceptable level
- By defining organizational policies that govern usage and outline control implementation expectations

Naturally, there is not a one-size-fits-all approach to security and Compliance, and as such, the risks, controls and policy outlined in this document serve as guidelines that should be modified and adapted to the precise needs of your organization.



Security Tiers

By virtue of an organization's usage of GenAI, amongst a plethora of other parameters (such as industry, product/service offering, Compliance obligations, sensitivity of data etc.), the level and type of risk exposure will differ.

As such, this framework categorizes its recommendations based on the following three Security Tiers:

ORGANIZATION TYPE

Security Tier 1

- **Employees use GenAI tools for general day-to-day tasks.**
- **Organization uses GenAI, such as a Large Language Model (LLM), to conduct business-critical processes and has dedicated resources to actively training the LLM.**
- **Organization has deployed GenAI within the production environment of the product/service offering.**

ORGANIZATION TYPE

Security Tier 2

- **Employees use GenAI tools for general day-to-day tasks.**
- **Organization uses GenAI, such as a LLM, to conduct business-critical processes and has dedicated resources to actively train the LLM.**

ORGANIZATION TYPE

Security Tier 3

- **Employees use GenAI tools for general day-to-day tasks.**

The GenAI Risk Register

ANEC-AI-1
Input of unsanitized PII/PHI

GenAI tools are commonly used for performing functions/analysis on large data sets. As such, it must be ensured that inputs are void of data that is subject to regulatory or contractual limitations, as to avoid infringement. Employees may knowingly input such data as the process of masking or anonymizing the data may be excessively strenuous. Alternatively, employees may unwittingly input such data due to a lack of awareness of what constitutes PII/PHI, or simply due to a lack of awareness of the data content. Regardless, formalized processes should be implemented to ensure employees understand their individual responsibilities and the broad implications of PII/PHI misuse. Employee access to PII/PHI should be restricted, based on 'least privilege'. Furthermore, employees accessing PII/PHI for training LLMs provided with an obscured data set (such as data that has been subjected to sanitization mechanisms).

RISK EVENT DESCRIPTION	Employee inputs privacy data (PI/PHI) that is subject to regulatory or contractual obligations. This privacy data is then used by the GenAI tool for learning purposes and outputted to other users. This constitutes a regulatory/contractual breach.
SECURITY TIER APPLICABILITY	ST1 ST2 ST3
RISK EFFECT	Confidentiality
RISK SOURCE	Human error.
THREAT	GenAI tool using user-inputted data for learning purposes and outputting the same data in response to user prompts.
VULNERABILITY	Undefined, unfollowed or misunderstood organizational policy relating to GenAI tool usage.
MITIGATING CONTROLS	1.1 Policy, 1.2 Policy augmentation, 2.1 Awareness, 3.1 Segregation, 4.1 Data sanitization, 4.2 Banner, 7.1 Input validation, 7.4 Sensitive data discovery.

ANEC-AI-2
Inaccurate information usage

The output generated by GenAI tools is delivered in a manner that is both highly efficient and highly convincing, and therefore employees (ST2-3) or end-users (ST1) are likely to take outputs at face value. However, in doing so, the organization is at risk of not only consuming inaccurate information, but also of perpetuating it. If the organization were to stand accused of spreading misinformation, this could have legal implications, cause significant reputational damage, and result in a loss of customer trust. Therefore, the organization should define a manual QA methodology for reviewing output data (such as a formal legal review, peer review, technical review or managerial review). Where GenAI is used in the production environment, a disclaimer should be presented to end-users regarding 3rd party responsibility for the accuracy of outputted information.

RISK EVENT DESCRIPTION	Inaccurate information outputted from GenAI tool resulting in loss of customer trust and/or reputational damage.
SECURITY TIER APPLICABILITY	ST1 ST2 ST3
RISK EFFECT	Integrity
RISK SOURCE	GenAI tool error.
THREAT	GenAI tool outputs inaccurate information.
VULNERABILITY	Lack of formalized data integrity/accuracy validation protocols or customer facing disclaimer.
MITIGATING CONTROLS	1.1 AI Policy, 2.1 Awareness, 3.2 Agreement, 5.1 Disclaimer, 5.2 Opt-out, 5.3 Consent, 5.4 Terms of Use, 7.2 Output validation, 7.5 Zero trust architecture.

ANEC-AI-3
Copyrighted information usage

Many LLMs draw data from an abundance of sources, some of which have been shown to contain copyrighted/proprietary information. Usage of this information or models trained/fine-tuned on copyrighted information would therefore constitute copyright infringement and potential legal action. As such, formal protocols should be implemented to validate the source of all data within prompt responses, datasets used for initial model training as well as datasets used to fine-tune trained models prior to usage (this may include a formal legal review).

RISK EVENT DESCRIPTION	Organizational usage of copyrighted information outputted from GenAI tool resulting in legal action against the organization.
SECURITY TIER APPLICABILITY	ST1 ST2 ST3
RISK EFFECT	
RISK SOURCE	GenAI tool error.
THREAT	GenAI tool outputs copyrighted information.
VULNERABILITY	Lack of formalized data source verification protocols.
MITIGATING CONTROLS	1.1 AI Policy, 1.2 Policy augmentation, 2.1 Awareness, 5.1 Disclaimer.

ANEC-AI-4
Input of sensitive business information

GenAI tools' ability to provide instant, valuable feedback on inputted data such as source code, financial data, architecture diagrams etc. may lead to an unintentional intellectual property leakage, in the event that the sensitive data is then presented to other users. As such, technical controls should be in place to ensure that GenAI usage is conducted in an environment that is segmented, where the tool does not utilize the input prompts for training any other models.

RISK EVENT DESCRIPTION	Employee inputs sensitive business information or intellectual property (such as source code, financial data, confidential diagrams etc.). This information is then used by the GenAI tool for learning purposes and may be outputted to other users. This would be deemed an IP data breach.
SECURITY TIER APPLICABILITY	ST1 ST2 ST3
RISK EFFECT	Confidentiality
RISK SOURCE	Human error.
THREAT	GenAI tool using user-inputted data for learning purposes and outputting the same data in response to user prompts.
VULNERABILITY	Undefined, unfollowed or misunderstood organizational policy relating to GenAI tool usage, or LLM that uses inputted data to train non-organization owned models.
MITIGATING CONTROLS	1.2 Policy augmentation, 3.1 Segregation, 4.2 Banner, 6.1 Training.

ANEC-AI-5
Vulnerable code usage

GenAI tools are capable of instantaneously generating code that may take developers significant time to write, and as such there are circumstances under which specific prompt types may provide crucial shortcuts. However, these 'shortcuts' may prove costly in the long run if secure software development best practices are not followed, such as performing SAST, peer review processes, and testing/QA prior to production deployment. Moreover, the aforementioned best practices should be monitored and enforced by leadership, creating a secure development culture.

RISK EVENT DESCRIPTION	Developer uses vulnerable GenAI outputted code which makes its way into the production environment.
SECURITY TIER APPLICABILITY	ST1 ST2
RISK EFFECT	Confidentiality Integrity Availability
RISK SOURCE	Human error.
THREAT	GenAI tool outputs code that is vulnerable to exploits.
VULNERABILITY	Lack of secure software development practices.
MITIGATING CONTROLS	1.2 Policy augmentation, 6.1 Training, 6.2 Threat modeling.

ANEC-AI-6
Advanced social engineering attack

Prior to GenAI tools, adversarial phishing/smishing attacks most often had tell-tale signs, allowing risk-aware recipients to raise a flag and not fall victim. In turn, awareness training provided to employees tends to focus on typos/unconvincing language/out-of-the-ordinary context. However, with the help of GenAI, adversaries can now overcome these shortcomings, and produce high-quality phishing/smishing attacks that are much more difficult to detect. As such, organizations must ensure that employee awareness training curricula shift focus to teaching more technical approaches to recognizing these attack vectors, as well as ensuring that there are sufficient technical safeguards in place to filter out these attacks before they reach employees.

RISK EVENT DESCRIPTION	Adversary uses high-quality GenAI outputs to create advanced phishing campaigns, successfully achieving a foothold in organizational systems.
SECURITY TIER APPLICABILITY	ST1 ST2 ST3
RISK EFFECT	Confidentiality Integrity Availability
RISK SOURCE	Malicious actor.
THREAT	GenAI tool outputs phishing content without expected tell-tale signs.
VULNERABILITY	Lack of employee ability/technical ability to detect malicious emails.
MITIGATING CONTROLS	2.2 Phishing, 7.3 Email filtering.

ANEC-AI-7
LLM dependency

GenAI LLMs possess data processing capabilities that are largely unparalleled, which may lead to excessive dependence, especially if they come in place of human expertise that is capable of performing the function in the event of LLM outage/downtime. If this business process directly ties to a contractual obligation (such as an SLA), this could pose significant risk to the organization. As such, effective disaster recovery and business continuity processes should be developed, and periodically tested.

RISK EVENT DESCRIPTION	Implementation of a GenAI LLM may result in excessive dependency. In the event of LLM outages/downtime the organization may lose a business-critical process that can no longer be handled by human expertise, which may in turn result in reputational or financial damage.
SECURITY TIER APPLICABILITY	ST1 ST2
RISK EFFECT	Availability
RISK SOURCE	Organizational process.
THREAT	GenAI tool replaces an organizational process and significantly exceeds previous workforce capabilities.
VULNERABILITY	Lack of effective BIA and DRP/BCP.
MITIGATING CONTROLS	8.1 Impact, 8.2 Recovery and continuity.

ANEC-AI-8
Biased output data

GenAI LLMs output data is based on training data that may stem from an unreliable or unverified source. As such, output data may possess the same biases or discriminations as the source data. This can include racial or gender discrimination, or any other protected characteristics, and can therefore, in turn, result in legal action against the organization, as well as reputational damage. This can be partly mitigated by the implementation of technical controls that ensure the diversity and integrity of training data, as well as an end-user facing disclaimer.

RISK EVENT DESCRIPTION	GenAI LLMs output data that is biased or outright discriminatory, which may result in reputational damage, or even legal action against the organization.
SECURITY TIER APPLICABILITY	ST1
RISK EFFECT	Integrity
RISK SOURCE	GenAI tool error.
THREAT	GenAI tool outputs biased or discriminatory data.
VULNERABILITY	Unvalidated/unverified training data.
MITIGATING CONTROLS	7.1 Input validation, 7.2 Output validation, 7.5 Zero trust architecture.

ANEC-AI-9
Prompt injection

GenAI LLMs can be configured to restrict output to specific parameters, using guardrails. However, adversaries are constantly seeking to design prompts that are capable of circumventing these guardrails, allowing them to utilize LLMs for potentially malicious activity. Many GenAI tools are currently releasing frequent updates to attempt to improve prompt injection resistance, and as such it should be ensured the most up-to-date version of the tool is used. Furthermore, guardrails should be put in place and continuously improved.

RISK EVENT DESCRIPTION	GenAI LLM is manipulated to accept prompts that do not conform to the configured parameters (guardrails), resulting in a cyber attack.
SECURITY TIER APPLICABILITY	ST1 ST2
RISK EFFECT	Confidentiality Integrity Availability
RISK SOURCE	GenAI tool error.
THREAT	GenAI tool is subjected to a prompt injection attack.
VULNERABILITY	Ineffective guardrails.
MITIGATING CONTROLS	7.6 Updates.

ANEC-AI-10
Direct LLM attack

Where GenAI LLMs are used to provide production environment functionality to end-users, adversaries may seek to conduct attacks that result in events such as a data breach, poisoned output data, denial of service and more. This may be achieved by directly attacking the LLMs back-end and gaining a foothold into the LLM management layer and data sets (e.g. through the source code or by accessing the GenAI tool itself). As such, maintaining good credential hygiene as well as ensuring that GenAI vendors adhere to industry-standard security practices should be ensured.

RISK EVENT DESCRIPTION	GenAI LLM tool used in production environment is breached via the back-end (e.g. through organization SCM tool or through the GenAI tool itself), resulting in a cyber attack.
SECURITY TIER APPLICABILITY	ST1
RISK EFFECT	Confidentiality, Integrity, Availability
RISK SOURCE	Malicious actor.
THREAT	GenAI tool is subjected to a direct attack.
VULNERABILITY	Lack of organizational credential hygiene or lack of vendor security protections.
MITIGATING CONTROLS	3.3 Assessment, 7.7 Credentials.

The anecdotes AI Framework

The following controls listing aims to provide proactive objectives for effective organizational measures in reducing AI risk exposure.

1. Governance

CONTROL NAME

1.1 AI Policy

CONTROL DESCRIPTION

The organization should develop an AI-specific policy, outlining usage limitations and expectations, and require all employees to read and formally acknowledge the policy prior to GenAI tool usage and development.

APPLICABILITY

ST1 ST2 ST3

EQUIVALENT CONTROLS

- ISO/IEC 27001 - ISMS 5.2
- CSA STAR A&A - 1
- NIST 800-53 REV 5 - AU-1

CONTROL NAME

1.2 Policy Augmentation

CONTROL DESCRIPTION

The organization should augment existing policies (such as the Acceptable Use policy, the SDLC policy etc.) to explicitly reference boundaries for AI usage and provide effective implementation guidance.

APPLICABILITY

ST1 ST2 ST3

EQUIVALENT CONTROLS

- ISO/IEC 27001 - ISMS 5.2
- CSA STAR A&A - 1
- NIST 800-53 REV 5 - AU-1

2. Training

CONTROL NAME

2.1 Awareness

CONTROL DESCRIPTION

The organization should augment the existing employee awareness training curriculum to include AI usage limitations, as stipulated in organizational policy.

APPLICABILITY

ST1 ST2 ST3

EQUIVALENT CONTROLS

- ISO/IEC 27001 - A.7.2.2
- CIS V8 - 14.1
- CSA STAR HRS - 11
- NIST 800-53 REV 5 - AT-2

CONTROL NAME

2.2 Phishing

CONTROL DESCRIPTION

The organization should augment the existing employee awareness training curriculum to address the recognition of more advanced adversarial tactics arising from GenAI tool usage.

APPLICABILITY

ST1 ST2 ST3

EQUIVALENT CONTROLS

- ISO/IEC 27001 - A.7.2.2
- CIS V8 - 14.2
- CSA STAR HRS - 11
- NIST 800-53 REV 5 - AT-2(3)

3. GenAI Vendors

CONTROL NAME

3.1 Segregation

CONTROL DESCRIPTION

The organization should ensure that where possible, a version of the GenAI tool is used that does not use inputted data to train or improve non-organization owned models (instance segregation).

APPLICABILITY

ST1 ST2 ST3

CONTROL NAME

3.2 Agreement

CONTROL DESCRIPTION

The organization should ensure that the GenAI vendor formally bears responsibility for the protection of inputted data, as well as liability for potential data inaccuracies.

APPLICABILITY

ST1 ST2 ST3

EQUIVALENT CONTROLS

- ISO/IEC 27001 - A.15.1.3
- CIS V8 - 15.4
- CSA STAR STA - 09

CONTROL NAME

3.3 Assessment

CONTROL DESCRIPTION

The organization should ensure that the GenAI vendor has undergone a security and privacy assessment (in accordance with the organizational Vendor Management policy) to ensure that the vendor adheres to industry-standard security practices.

APPLICABILITY

ST1 ST2

EQUIVALENT CONTROLS

- ISO/IEC 27001 - A.15.2.1
- CIS V8 - 15.5
- CSA STAR STA - 13
- NIST 800-53 REV 5 - SR-3

4. Privacy

CONTROL NAME

4.1 Data Sanitization

CONTROL DESCRIPTION

The organization should ensure that training data has been sanitized (such as by masking), especially when the data set is derived from bulk source data.

APPLICABILITY

ST2 ST3

EQUIVALENT CONTROLS

ISO/IEC 27701 - 6.11.3.1

CONTROL NAME

4.2 Banner

CONTROL DESCRIPTION

The organization should ensure that where feasible, banners are presented to users near input fields, reminding them of usage policies.

APPLICABILITY

ST1 ST2

EQUIVALENT CONTROLS

NIST 800-53 REV 5 - AC-8

5. Customer Obligations

CONTROL NAME

5.1 Disclaimer

CONTROL DESCRIPTION

The organization should ensure that customers are presented with a disclaimer prior to usage of GenAI elements within the product offering, which recommends input data practices and warns of potential output data inaccuracies, and defers responsibility for outputted data to the GenAI vendor.

APPLICABILITY

ST1

CONTROL NAME	CONTROL DESCRIPTION	APPLICABILITY
5.2 Opt-out	The organization should ensure that customers are able to opt-out of GenAI elements within their product offering, and are informed of GenAI usage/functionality when it is enabled by default.	ST1

CONTROL NAME	CONTROL DESCRIPTION	APPLICABILITY
5.3 Consent	The organization should ensure that customers formally consent to GenAI usage within the product offering.	ST1

CONTROL NAME	CONTROL DESCRIPTION	APPLICABILITY
5.4 Terms of Use	The organization should update the Terms of Use to reflect GenAI usage within the product offering and make existing customers aware of the changes retroactively.	ST1

6. Secure Development

CONTROL NAME	CONTROL DESCRIPTION	APPLICABILITY
6.1 Training	The organization should ensure that all relevant employees undergo training in AI secure development practices (as per the augmented SDLC policy) prior to using GenAI tools for development purposes.	ST1 ST2 EQUIVALENT CONTROLS ISO/IEC 27001 - 14.2.5 CIS V8 - 14.9 CSA STAR AIS - 04 NIST 800-53 REV 5 - AT-3

CONTROL NAME

6.2 Threat Modeling

CONTROL DESCRIPTION

The organization should conduct threat modeling as part of the development process of GenAI product functionality.

APPLICABILITY

ST1 ST2

EQUIVALENT CONTROLS

CIS V8 - 16.14

NIST 800-53 REV 5 - SA-11(2)

7. Technological Protection

CONTROL NAME

7.1 Input Validation

CONTROL DESCRIPTION

The organization should ensure that acceptable input parameters are defined and tested and are enforced by the GenAI tool.

APPLICABILITY

ST1 ST2

EQUIVALENT CONTROLS

NIST 800-53 REV 5 - SI-10

CONTROL NAME

7.2 Output Validation

CONTROL DESCRIPTION

The organization should implement an output validation process for LLMs in order to ensure that they continue to meet defined accuracy and non-bias criteria.

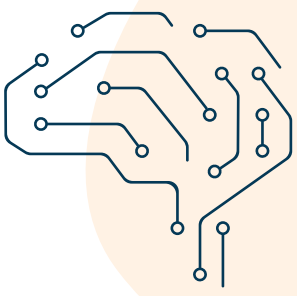
APPLICABILITY

ST1 ST2

EQUIVALENT CONTROLS

CSA STAR CCC - 02

NIST 800-53 REV 5 - SI-15



CONTROL NAME

7.3 Email Filtering

CONTROL DESCRIPTION

The organization should implement email filtering technologies (such as DMARC) to prevent potentially malicious emails from reaching employee inboxes.

APPLICABILITY

ST1 ST2 ST3

EQUIVALENT CONTROLS

CIS V8 - 9.5, 9.7

CONTROL NAME

7.4 Sensitive Data Discovery

CONTROL DESCRIPTION

The organization should implement a sensitive data discovery mechanism to ensure data repository sanitization.

APPLICABILITY

ST1 ST2

EQUIVALENT CONTROLS

CSA STAR DS - 03

NIST 800-53 REV 5 - AC-4(25)

CONTROL NAME

7.5 Zero Trust Architecture

CONTROL DESCRIPTION

The organization should implement a zero trust architecture to prevent unauthorized access to sensitive organizational assets (such as DBs containing PII/PHI, or LLMs).

APPLICABILITY

ST1 ST2

EQUIVALENT CONTROLS

CSA STAR IAM - 05

NIST 800-53 REV 5 - AC-6

CONTROL NAME

7.6 Updates

CONTROL DESCRIPTION

The organization should ensure that the GenAI tool used is the most up-to-date and secure version.

APPLICABILITY

ST1 ST2 ST3

EQUIVALENT CONTROLS

CIS V8 - 12.1

NIST 800-53 REV 5 - SI-2(4)

CONTROL NAME

7.7 Credentials

CONTROL DESCRIPTION

The organization should enforce user credentials (such as passwords and access keys) that are periodically rotated and meet industry-standard strength. Multi-factor authentication should be enforced where possible.

APPLICABILITY

ST1 ST2

EQUIVALENT CONTROLS

- ISO/IEC 27001 - A.9.3.1
- CIS V8 - 6.3
- CSA STAR IAM - 15
- NIST 800-53 REV 5 - IA-5(1)

8. Dependency Resilience

CONTROL NAME

8.1 Impact

CONTROL DESCRIPTION

The organization should perform business impact analysis (BIA) for all GenAI tool usage.

APPLICABILITY

ST1 ST2

EQUIVALENT CONTROLS

- ISO/IEC 27001 - 17.1.1, 17.1.2, 17.1.3
- CSA STAR BCR - 02

CONTROL NAME

8.2 Recovery and Continuity

CONTROL DESCRIPTION

The organization should develop clearly defined disaster recovery and business continuity plans for any GenAI tool usage that has been deemed to be business critical. These plans should be tested periodically.

APPLICABILITY

ST1 ST2

EQUIVALENT CONTROLS

- ISO/IEC 27001 - 17.1.2, 17.1.3
- CSA STAR BCR - 03, BCR - 04, BCR - 09
- NIST 800-53 REV 5 - CP-2, CP-4

Generative AI Implementation Policy Template

Purpose

[Org. name] is committed to preserving organizational data confidentiality, integrity, and availability and in turn adhering to contractual and regulatory obligations. As such, [Org. name] defines and outlines this organizational Generative Artificial Intelligence (henceforth referred to as "GenAI") policy, in order to uphold these obligations by establishing a governance baseline, upon which technological controls can be built.

Scope

This policy applies to [Org. name] employees in all offices, whether working in the office or remotely, and whether employed in a full or part-time capacity. This policy applies to GenAI usage across two different areas of application:
1. Employee usage of GenAI tools for day-to-day activities (for example, inputting prompts to a service such as ChatGPT).
2. [Employee usage of GenAI tools, specifically Large Language Models (henceforth referred to as "LLMs") for business-critical processes.]

[Certain requirements outlined within this policy that are relevant to the organization as a whole and additional GenAI requirements not outlined within this policy that are more department-specific are further detailed in [Org. name] policies, which have been augmented in order to accommodate GenAI tool usage. These policies are:

- [Acceptable Use policy]
- [Secure Development Lifecycle policy]
- [Security Awareness and Training policy]
- [Vendor Management policy]
- [Disaster Recovery and Business Continuity policy]

For organization-wide requirements, this policy provides a brief overview and then makes reference to the topic-specific policy.]

Policy life-cycle

This policy is made readily available to all [Org. name] employees via the [internal portal/shared drive/HR tool] and is actively communicated to all employees upon first release, as part of the onboarding process and annually thereafter.

This policy is made available to any interested parties upon request, as deemed appropriate by the [CISO] [and the DPO].

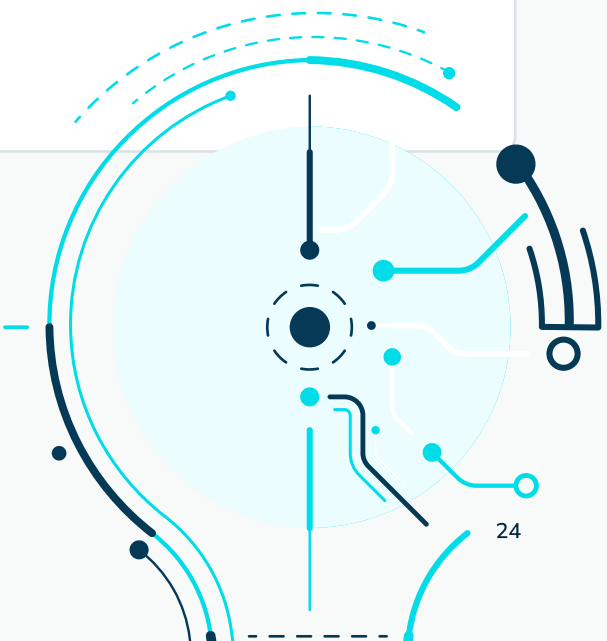
This policy is reviewed and formally approved by the [CISO] [and the DPO] on an annual basis or following any changes.

Background

[Org. name] acknowledges that GenAI tools represent significant opportunities for increased efficiency and productivity, and as such does not wish to prevent or discourage employees from their usage. However, [Org. name] also acknowledges that GenAI tools present new, significant risks to the organization, that without the implementation of effective mitigating controls, will exceed the organizational risk appetite.

As such, [Org. name] has designed a control set that aims to reduce the risks associated with GenAI usage to an acceptable level. These control areas include:

- Defining acceptable use of GenAI tools
- [Implementing additional awareness training content]
- [GenAI vendor security and privacy requirements]
- [Customer facing obligations]
- [Data input/output validation]
- [Dependency resilience]



Acceptable use

Employees using GenAI tools shall conform to the behavioral expectations laid out in the organizational [Code of Conduct], in particular pertaining to refraining from the input of profanity or of discriminatory content or views.

Furthermore, employees shall refrain from:

- Creating prompts that contain organizational intellectual property (such as [source code, financial information, trade secrets etc.]).
- Creating prompts that contain any data that is subject to regulatory or compliance limitations, such as personal identifiable information (PII) or protected health information (PHI).
- Using GenAI tool outputs without an effective accuracy validation process.
- Publishing GenAI tool outputs without a [Legal dept.] approved disclaimer.

In the event that exceptions to the aforementioned are required in order to support a business process, an exception request shall be made in writing and formally approved by the [CISO] [and the DPO].

The [Org.name] [Acceptable use policy] has been appended to include the aforementioned requirements.

Employee training

Employees using GenAI tools shall undergo additional information security awareness training that specifically includes:

- Acceptable use of GenAI tools as outlined above.
- Recognizing advanced social engineering attacks.

Engineering teams using GenAI tools shall undergo additional training that specifically includes:

- Risks associated with usage of code outputted by GenAI tools.
- Static Application Security Testing (SAST) process requirements.
- Peer review requirements.

The [Org.name] [Security awareness and training policy] and [Secure development lifecycle policy] have been appended to include and further detail the aforementioned requirements.

Vendor management

Prior to permitting the organization-wide usage of a GenAI tool, the [GRC team] shall perform a vendor risk assessment in accordance with the [Vendor management policy] and define minimum security requirements to approve the tool.

As a baseline, the requirements of GenAI tools are:

- A version offering that does not use inputted prompts to train or improve shared models, allowing for a segregated user instance.
- A formal acknowledgement of responsibility for data protection and liability for potential data inaccuracy.

The [Org.name] [Vendor management policy] has been appended to include and further detail the aforementioned requirements.

[LLM dependency]

[Org. name] shall perform business impact analysis for each trained LLM.

When using LLMs for processes deemed critical as per the business impact analysis, [Org. name] will ensure that excessive dependency on the LLM does not occur, as to avoid excessive risk exposure in the event of LLM downtime/failure. This shall be achieved by:

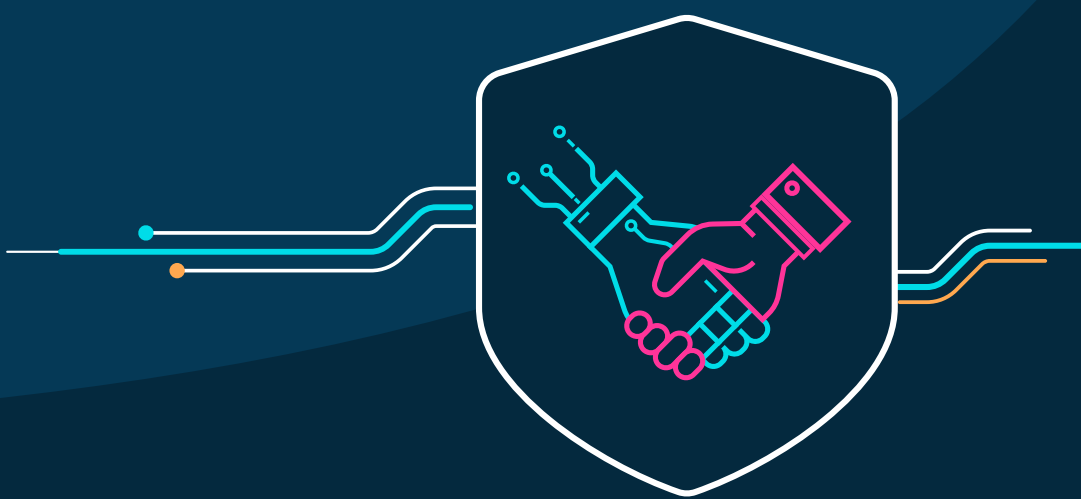
- The development of [Disaster recovery and Business Continuity plans] that account for business-critical LLM downtime/failure.
- The periodic testing of these [Disaster recovery and Business Continuity plans] to ensure their effectiveness.
- The implementation of an employee [Professional development program] as to ensure that the introduction of LLMs does not come at the expense of employee expertise and competence.

The [Org.name] [Disaster recovery and Business Continuity plans] and [Professional development program] have been appended to include and further detail the aforementioned requirements.

Summary

As we conclude the creation of this document, the continued innovation in Generative AI remains at the front and center of not only the Security and Compliance ecosystem, but of global news as well. And as this is unlikely to change in the near future, new use cases will surely raise questions around new, unintended risks and threats.

It is imperative that organizations stay abreast of new developments, and implement effective mitigation strategies. anecdotes remains committed to providing the community with practical solutions and tools and will continue to evolve this framework to address new challenges (so stay tuned for the next iteration!).



anecdotes

anecdotes is the leading technology provider for Compliance leaders. Powered by data, the anecdotes Compliance Operating System (OS) transforms security Compliance from a box-ticking exercise into a powerful driver of growth. With a variety of applications powered by verified data, Compliance leaders as well as advisory firms and auditors, can turn manual, time-consuming, and siloed tasks into an automated, continuous, and strategic Compliance program. For more information, visit anecdotes.ai.