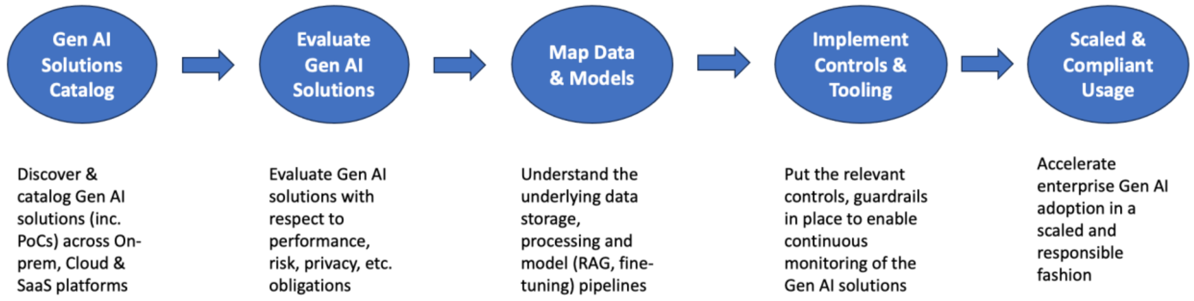
Generative AI Governance Checklist



Discover & catalog Gen AI solutions (inc. PoCs) across On-prem, Cloud & SaaS platforms

Evaluate Gen AI solutions with respect to performance, risk, privacy, etc. obligations

Understand the underlying data storage, processing and model (RAG, fine-tuning) pipelines

Put the relevant controls, guardrails in place to enable continuous monitoring of the Gen AI solutions

Accelerate enterprise Gen AI adoption in a scaled and responsible fashion

Fig: Gen AI Governance

The document is structured in the form of a Generative AI (Gen AI) Solution Card. It aims to provide the necessary details of the underlying Gen AI solution, such that it can be deployed in a scalable and responsible fashion, complying with applicable regulations. Given this, it serves as an initial governance checklist for the Legal, Privacy, Auditing, Responsible AI, Ethical AI, etc. teams of the enterprise to understand the design, capabilities, and constraints of the Gen AI solution; who can then decide on the further processing steps for their respective streams (as needed). This shows a proactive and responsible approach by the enterprise collaborating with community at large to deploy state-of-the-art technology solutions that benefit society as a whole.

Question	References	Status
Application description		
Provide Use-case details: What business problem does it solve? Its capabilities. What other techniques were considered before deciding to use Gen AI to solve this problem?		
Business Unit(s) involved		
Business Unit (s) impacted		
Describe Gen AI solution: provide details of LLMs, SLMs used, solution architecture, e.g. RAGs, and also indicate if fine-tuning is involved. *Only whitelisted LLMs and SLMs are allowed by default. Usage of other LLMs and SLMs will require additional reviews.	Enterprise LLM classification (link) Gen AI Architecture Patterns (link)	
List Input and Output data types		
Evaluation Strategy		
Gen AI solution evaluation details: What types of tests have been performed? How does the solution perform with respect to Gen AI solution metrics, e.g., correctness, groundedness, contextual relevance, etc.	Enterprise Use-case based Evaluation of LLMs (link)	

Monitor Gen AI solution Are the inputs and output logged? What is the feedback and review frequency? How long would it take to react to an undesired behavior?		
Onboard end-users Have the users been trained on the appropriate usage of the Gen AI solution? Is documentation readily available to all users?		
Address infra and cybersecurity risks Provide details of the cloud provider, landing zone, secure deployment architecture. Describe measures undertaken to address LLM specific security risks, e.g., prompt injection, jailbreaks and adversarial attacks.	Cloud Usage Guidelines (link) Information Security Policy (link)	
Human oversight Describe the human intervention points in the Gen AI solution development and deployment lifecycle.		
Responsible AI		
Transparency Provide details of how the Gen AI solution notifies the end-users that the responses are generated by Gen AI. Outline mechanisms in place for users to provide feedback and the underlying process to incorporate them into the Gen AI solution.	Gen AI Design Principles (link)	
Bias & Discrimination Describe measures undertaken to address bias in training data (including RAGs and fine-tuning) and guardrails to ensure non-discriminative responses.	Responsible LLMOps	
Risk level Does the Gen AI solution provide responses that impact access to, or approval for, housing or accommodations, education, employment, credit, healthcare, or criminal justice? Specify the 'risk level' of the Gen AI solution according to Enterprise AI Risk Management Policy. Describe the process and tooling in place to monitor that the determined Gen AI solution's risk level will or has not changed over time (e.g., changes with respect to users, data, privacy, cost). EU AI Act specific Does the Gen AI solution display significant generality to be deployed as an independent AI model that can be integrated into downstream applications? (e.g., pre-trained LLMs, fine-tuned SLMs). Specify the 'risk level' of the Gen AI solution according to the EU AI Act.	Enterprise AI Risk Management Policy (link) EU AI Act Compliance Checker (link)	
Toxicity List details of guardrails in place with respect to both input prompts and responses generated by the Gen AI solution – to address concerns with respect to toxicity, illicit content, etc.		

Explainability Provide high-level business process logic of the Gen AI solution. Is it possible to explain the inner workings of the Gen AI solution to the end-user (including that of the underlying LLM)? – linking specific responses to the source data / documents. Outline any XAI model used with respect to local and global explainability.		
Privacy – determine need for DPIA		
Data sensitivity Provide details of confidential and sensitive data processed by the Gen AI solution. Does it handle personal data?		
Data processing Outline legal basis to process personal and sensitive data. Measures to ensure that the processing logic has not changed from the original purpose (for which consent was obtained), and is deleted after the stipulated period.		
Access control Describe the access control process. Measures to ensure that access to both training and conversational data (logs) is only available on a need-to-know basis.		
Privacy protection Provide details of the privacy preserving mechanisms in place to address risks specific to the Gen AI solution, with respect to both training and conversational data (logs). Outline measures to ensure that the Gen AI solution does not leak private or sensitive data, esp. if the Gen AI solution is subject to adversarial attacks.	Generative AI Privacy Risks (link)	
Data Compliance Provide details of measures in place to comply with data subject requests with respect to access, objection, deletion, etc.		
Vendor Management		
3rd party involvement Specify details of any 3rd party involvement in design, development, deployment, and support of the Gen AI solution.		
LLM provider LLM vendor is a special type of 3rd party vendor in this case providing the underlying pre-trained LLM. Ensure that we opt-out of any data collection and model training features that will be used to (re-)train the underlying LLM.		
Data ownership Provide details of data ownership negotiated with the 3rd party (and LLM) vendor.		

Ensure that this includes both training data and conversation logs, as they are key to improving the Gen AI performance over time.		
Liability Provide details of the liability agreement with the 3rd party vendor. Focus on copyright infringement and intellectual property (IP) compliance inc. Open-source licenses.	Open-Source licensing primer for Enterprise AI/ML (link)	