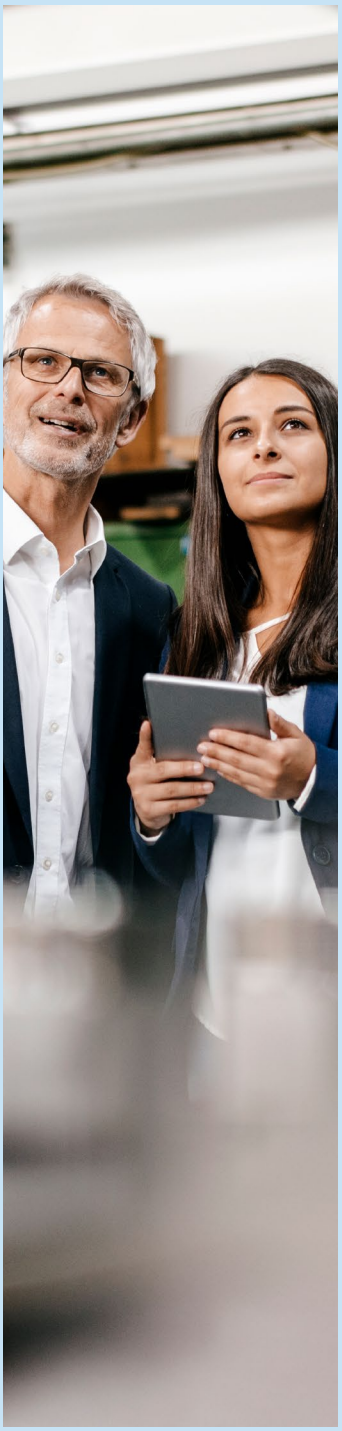
VIEW POINT



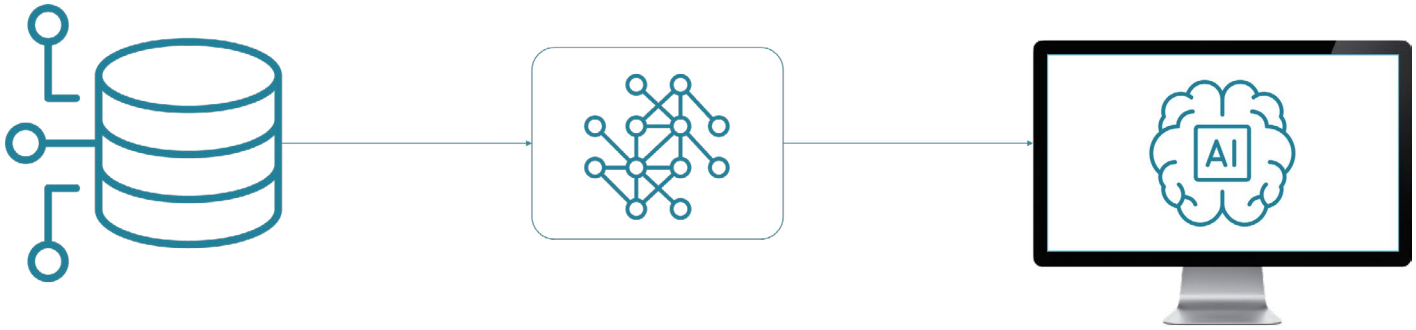
PROTECTION AND PRIVACY OVER PROFIT - BUILDING GUARD RAILS FOR GENERATIVE AI

Infosys®
Navigate your next

Enterprise Impact of Generative AI

The idea of a computer generating dynamic human-like content is exciting. Generative AI has been in existence for sometime, but the possibility of putting ChatGPT, AlphaCode, DALL-E-2 and other AI tools in the hands of every person on the planet through "co-pilot" software products holds great promise. According to Microsoft's 2023 Work Trend Index, 3 out of 4 professionals were open to delegate to AI to reduce their workload. If implemented responsibly, generative AI has the potential to create a ripple effect, transform the industries, accelerate innovation, and amplify human potential.

Generative AI refers to a family of AI algorithms which learn the representation or pattern of an artifact through data. Based on what they have learnt, they generate synthetic data or artifacts that preserve similarity to the input fed to them.



Large amount of unlabeled data

Transformer Algorithm which enables learning

Generative AI Applications

Figure 1 - Impact of data on Generative AI applications

"Generative" refers to the fact that these tools can find patterns across enormous sets of data and generate new content. Creativity has always been a uniquely human trait. Their most striking advance is in natural language capabilities, which are needed for many work activities. While ChatGPT is focused on text, other AI systems from major platforms can create audio, video, and images. Although generative AI is still in the initial stages, the potential applications for businesses are significant and wide-ranging. Generative AI could be used to create code, design products, create marketing content and strategies, provide customer service via chatbots, streamline the operations, analyze legal documents, and even speed up the scientific discoveries. It can be used on its own or with "humans in the loop"; the latter is more seemingly at present, given its prevailing level of maturity.

Generative AI tools are built on top of a variety of complex machine learning models which understand language and constantly keep learning. This learning exercise happens by training the algorithm through an exceptionally large volume of unlabeled data. For instance, ChatGPT, which is one of the most popular Generative AI applications, needed 570 GB of training data. Synthetic image generation algorithm like Stable diffusion AI consumed around 200 million to start generating meaningful output. To feed this voracious need for data, AI developers indiscriminately collect data without consideration of basic privacy principles such as data minimization or consent of individuals.

How can Generative AI cause privacy risks?

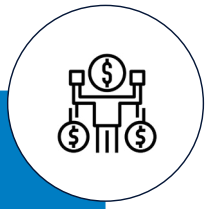
Generative AI is versatile and accessible. This explains why ChatGPT took only 5 days to reach a million users compared to Twitter which took 24 months to reach this marked milestone. These benefits come with an overshadow of training of Generative AI through an opaque process. This makes it a prime candidate for privacy risks, copyright infringement, potential for misuse and hallucinations.

Privacy Risks

Generative AI algorithms often require large amounts of data, and this data can include sensitive information about individuals. The process of training these models may not always adequately protect user privacy, leading to potential data breaches or unauthorized access.

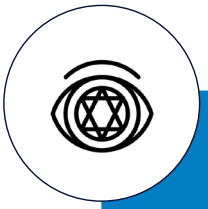
Copyright Infringement

As Generative AI can generate content that closely resembles existing works, there is a risk of copyright infringement. This could lead to legal issues if the AI-generated content is used without proper authorization from the original creators.



Potential for Misuse

While Generative AI has many practical applications, it also has the potential for misuse in creating fake news, deepfake videos, or other forms of disinformation, which may have harmful outcomes for individuals and society.



Hallucinations

Generative AI models may sometimes produce outputs that are unrealistic or nonsensical, known as "hallucinations." These erroneous outputs could mislead users or lead to inaccurate decision-making, if not properly controlled

No consent and opaque data collection

Collection of personal data through web-scraping may be risky without proper consent. Further multiple data sets combined causes the privacy risk to increase manifold times

Exploit User's digital trail

Many GA tools require user to login for access and retain user contact information, IP addresses, inputs and outputs of conversations the users have within and with the apps. Lack of data disposal can amplify the impact of data proliferation and violate storage limitation principle of privacy

Attack on AI model

Generative AI tools can be subjected to model inversion attacks. An attacker can use a generative AI tool to infer private information about the individual by observing the model's outputs.

Unauthorized access

Training data used for generative AI models if not adequately protected can be subjected to unauthorized access.

Figure 2 : Privacy risks of Generative AI

Building Guard Rails – How can enterprises get the best of generative AI without privacy and security risks?

The excuse of collecting data without any privacy guard rails in the name of innovation and staying ahead of competition is harmful to the state of data privacy and security. Generative AI tools may inadvertently share personal information about someone or may include an element such as a photo shared in social media. Enterprises might have trade secrets shared in training sets causing potential loss. Individuals may not even imagine how AI tools use personal data shared on social media and websites. Furthermore, they may not be able to remove their data from generative AI tools.

Risks of generative AI applications must be addressed across each layer where the application interacts with enterprise stakeholders. In our view, there are 3 key layers where privacy and protection controls can be built into generative AI.

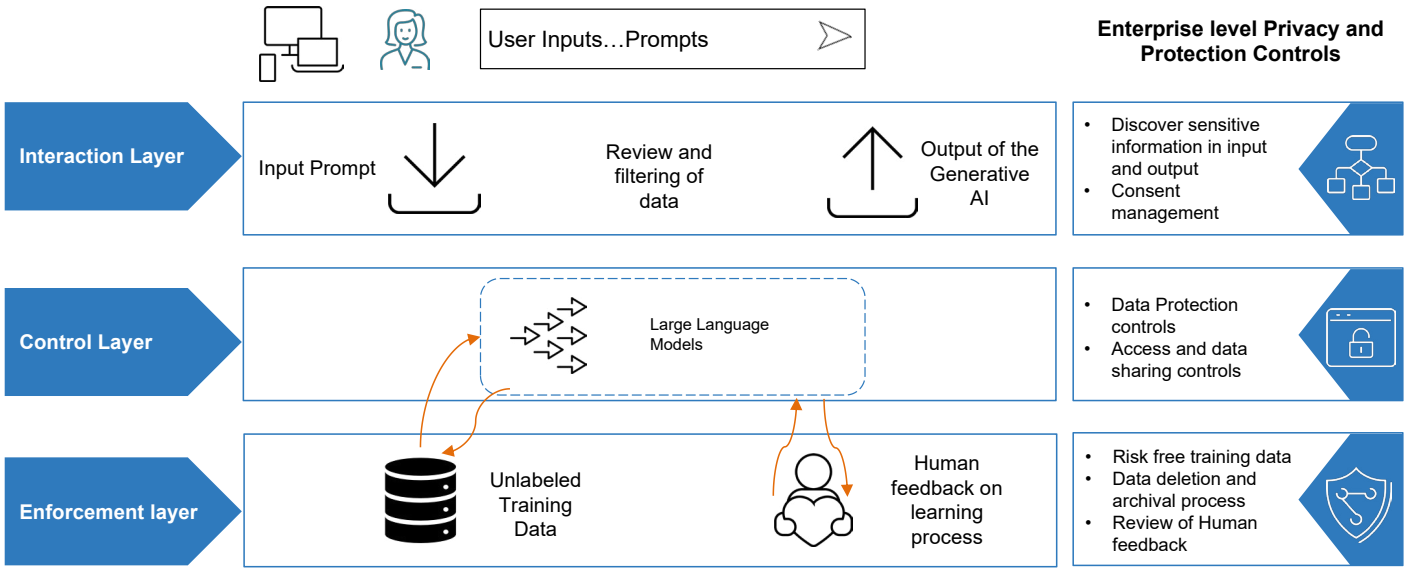


Figure 3 : Data Protection and Privacy controls for Generative AI

Layer	Privacy and Protection Controls	Impact
Interaction	• Discover sensitive information shared as input and in the output generated by the algorithm. Check for the security and privacy guidelines of the enterprise. • Check for consent of the data shared with the generative AI model.	This is the layer which has direct interaction with the user. The inputs from the user could have privacy risk to the model. The output to the user should also be checked for any breach of personal data or security risk. Furthermore, there is well designed consent governance to ensure there is no loss of personal data.
Control	• Check if the right data protection and access controls are in place for the model. • Check for data sharing controls.	Unauthorized access to models and unauthorizeddata sharing is avoided.
Enforcement	• Synthetic data for anonymized and bias test-ed data to be used as training data. • Check for data deletion and archival process after training. • Process review of the Human feedback of the learning process. • Conduct regular privacy impact assessment to identify privacy risk and plan mitigations.	Training data is checked for security and privacy risk.

Our recommendation for enterprise class Generative AI applications

The organizations exploring generative AI either expect or are already experiencing tangible results. Microsoft Azure OpenAI platform or AWS Code Whisperer can generate code from English prompts boosting employees' productivity. Enterprises can use Infosys 4D framework to build privacy controls for adoption of generative AI.

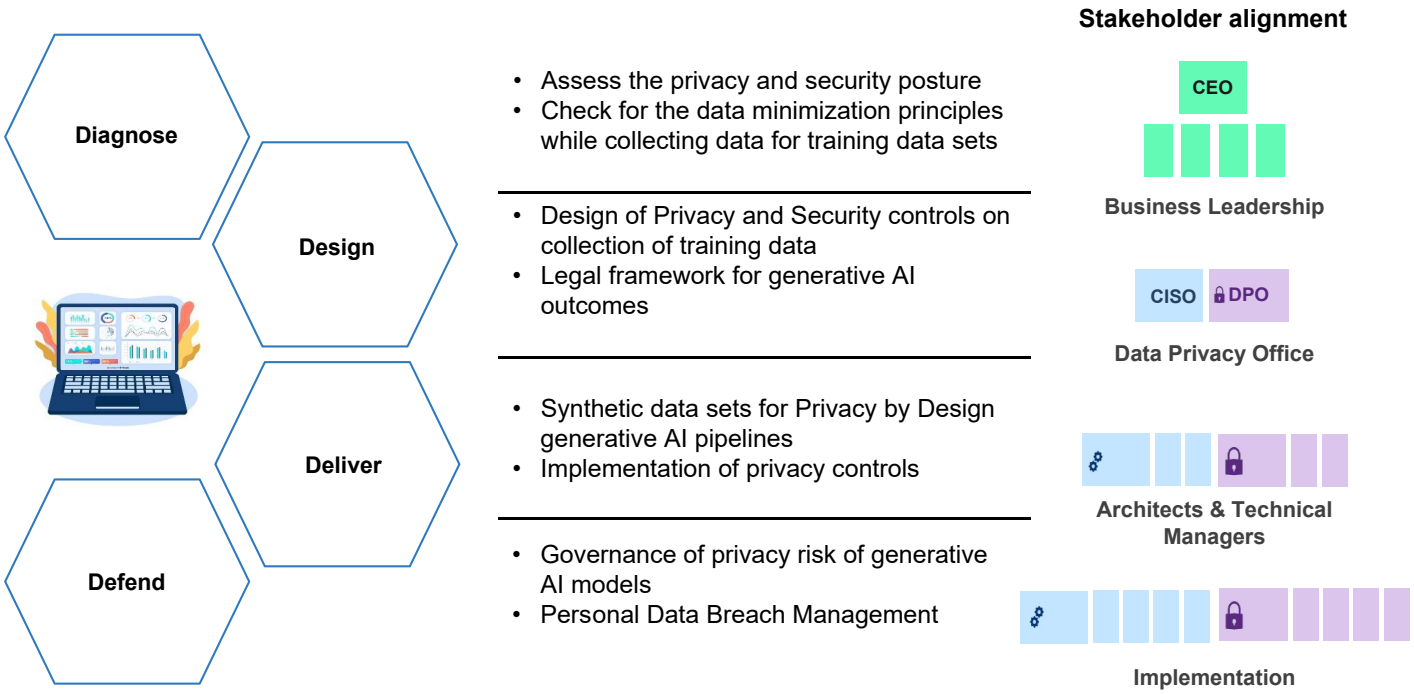


Figure 4 - Infosys 4D framework to accelerate enterprise scale adoption of Generative AI

The key assurance of generative AI is to streamline any routine language- or process-driven task virtually, supporting the human capabilities while freeing up more creative and productive uses of time. MakeMyTrip, a leading travel app company has introduced voice-aided booking in Indian languages to complement the work done by its human agents, considering the requirements of the nation which is home to over 400 native languages. With privacy, ethical and sustainable guard rails, enterprises can build scalable generative AI apps to rewrite complex workflows and engage with customers, partners, and employees in a better way; without breaching enterprise security and privacy.

References

Generative AI and the future of work in America - McKinsey Global Institute, July 2023

Beyond the Hype: Enterprise Impact of ChatGPT & Generative AI – Gartner, May 2023.

About the Author

Karthik Nagarajan

Senior Industry Principal

Karthik heads Infosys Data Protection and Privacy services. He has 17+ years of experience in product design and consulting services, with an expertise in AI, data privacy and customer experience strategy.

For more information, contact askus@infosys.com

Infosys®
Navigate your next