THE CAPCO INSTITUTE

JOURNAL

OF FINANCIAL TRANSFORMATION

TECHNOLOGICAL

Applied generative AI governance: A viable model through control automation

GERHARDT SCRIVEN | MARCEL BRAGA
DIOGO SANTOS | DIEGO SARAI

ARTIFICIAL INTELLIGENCE

#58 NOVEMBER 2023

a wipro company

THE CAPCO INSTITUTE

JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

CONTENTS

TECHNOLOGICAL

OPERATIONAL

ORGANIZATIONAL

DEAR READER,

As the financial services industry continues to embrace transformation, advanced artificial intelligence models are already being utilized to drive superior customer experience, provide high-speed data analysis that generates meaningful insights, and to improve efficiency and cost-effectiveness.

Generative AI has made a significant early impact on the financial sector, and there is much more to come. The highly regulated nature of our industry, and the importance of data management mean that the huge potential of AI must be harnessed effectively – and safely. Solutions will need to address existing pain points – from knowledge management to software development and regulatory compliance – while also ensuring institutions can experiment and learn from GenAI.

This edition of the Capco Journal of Financial Transformation examines practical applications of AI across our industry, including banking and fintechs, asset management, investment advice, credit rating, software development and financial ecosystems. Contributions to this edition come from engineers, researchers, scientists, and business executives working at the leading edge of AI, as well as the subject matter experts here at Capco, who are developing innovative AI-powered solutions for our clients.

To realize the full benefits of artificial intelligence, business leaders need to have a robust AI governance model in place, that meets the needs of their organizations while mitigating the risks of new technology to trust, accuracy, fairness, inclusivity, and intellectual property. A new generation of software developers who place AI at the heart of their approach is also emerging. Both GenAI governance and these 'Developers 3.0' are examined in this edition.

This year Capco is celebrating its 25th anniversary, and our mission remains as clear today as a quarter century ago: to simplify complexity for our clients, leveraging disruptive thinking to deliver lasting change for our clients and their customers. By showcasing the very best industry expertise, independent thinking and strategic insight, our Journal is our commitment to bold transformation and looking beyond the status quo. I hope you find the latest edition to be timely and informative.

Thank you to all our contributors and readers.

Lance Levy, **Capco CEO**

APPLIED GENERATIVE AI GOVERNANCE: A VIABLE MODEL THROUGH CONTROL AUTOMATION

GERHARDT SCRIVEN | Managing Principal, Capco[1]
MARCEL BRAGA | Principal Consultant, Capco
DIOGO SANTOS | Principal Consultant, Capco
DIEGO SARAI | Managing Principal, Capco

ABSTRACT

Generative AI has the potential to revolutionize the banking industry with hyper-personalization and advanced chatbots. However, the technology also poses risks to trust, accuracy, fairness, intellectual property, and confidentiality that all need to be mitigated to ensure that the benefits of Generative AI are realized. In this article, we explore practical considerations to help mitigate these risks through the construction of a governance framework that has a focus on AI explainability, intellectual property protection, and minimizing model hallucination. We then derive a control framework against these key outcomes and present technology solutions we built around automating some of the key controls towards making our governance model viable. Finally, we explore what other institutions are doing in the field of generative AI governance and discuss new emerging roles needed to execute against the governance model. In terms of practical application, we recommend that financial institutions start small when it comes to generative AI governance and focus on defining a "minimum governance model" on a use case by use case basis to minimize the time and cost footprint of governance. We also recommend that governance is implemented very early in the solution lifecycle so that it is baked in at root-level; hence, reducing churn and rework of the solution when industrializing the use case within the financial institution.

1. INTRODUCTION

Generative AI has the potential to revolutionize the banking industry, from a business as well as a technology perspective, by enabling hyper-personalization around financial planning, investment portfolios, product recommendations, and financial education. Moreover, personalized customer service can be provided to clients using the technology through advanced chatbots that provide tailored responses based on the customer's financial history and preferences.

However, as AI systems become more advanced and integrated into the banking industry, there is a growing need to understand and manage AI-related risks to ensure that the benefits of AI are realized while potential negative consequences are minimized.

More precisely, generative AI, which encompasses techniques such as deep learning and generative adversarial networks and include "large language models" (LLMs – generative AI that specializes in text understanding and generation), has the potential to create highly realistic and sophisticated outputs, including fake information and malicious code. This poses a range of risks, such as erosion of trust in financial institutions and the risk that AI may provide sub-par or incorrect recommendations and advice to bank personnel or the financial institution's customers.

Additional generative AI risks that financial institutions need to be aware of, and mitigate, include:

[1] We would like to thank the sponsors of this work: Alessandro Corsi and Luciano Sobral.

- **Bias and fairness:** generative AI models can inherit and perpetuate biases present in their training data, leading to biased content generation and reinforcing existing inequalities.
- **Intellectual property infringement:** generative AI models can generate content that infringes upon copyrights and trademarks, posing legal challenges.
- **Data protection:** to obtain the best results from generative AI for specialized tasks, it is often necessary to finetune the AI models with contextual information pertaining to the knowledge domain the solution will address, either in the form of training or via prompt engineering (crafting input instructions or queries to achieve desired outcomes when using Gen-AI). Herein lies an additional potential risk, that of protecting corporate intellectual capital as well as personal information of customers. For the latter, generative AI systems that process personal data must be designed in a way that protects the privacy of that data. This includes implementing appropriate security measures and providing individuals with control over their personal data, as stipulated by regulations such as the European Union's General Data Protection Regulation (GDPR).

As a secondary driver for governance: as one builds out automated solutions around addressing some of these risks, one also needs to be sure that one can trust the automated processes.

To create a holistic approach for managing AI risk, Tan (2023) presents a "generic AI risk management framework",[2] which consists of six pillars, of which "governance & oversight" is a key component to manage the other five pillars.

In this paper, we will explore the governance pillar in more depth and focus on the practical considerations (applied AI governance, a corresponding control framework, and emerging roles needed to manage generative AI) in order to help mitigate AI-related risks. The topics covered here will be particularly relevant for readers who are relatively new to implementing solutions using generative AI technology in corporate environments.

Figure 1: A generic AI risk management framework

Source: Tan (2023)

2. DEFINING A GENERATIVE AI GOVERNANCE MODEL

In establishing any governance model, a good starting point is to define the desired outcomes that one wants to achieve through applying the model. In the case of generative AI, there are three key outcomes that need to be considered.[3]

2.1 Being able to explain the results from AI

It is critical to be able to explain how AI, and in particular generative AI, arrived at a certain result.

- **Transparency:** explainability allows stakeholders to understand and trace how the AI system arrived at its conclusions or generated its outputs. This helps build trust and confidence in the technology.
- **Bias detection and mitigation:** explainability enables the identification of biases or unfairness in the AI system's outputs. By understanding the underlying processes and decision making, biases can be detected and addressed, leading to fairer and more equitable outcomes.

[2] https://tinyurl.com/27898j48

[3] It should be noted that the three outcomes discussed below are not exhaustive and that there are other dimensions of generative AI governance. Others include protecting AI models from adversarial attacks and ensuring that AI models are performant and scalable. However, these are well established AI-related governance topics, whereas the three key items listed below require new or significant additional thinking specifically for generative AI.

- **Error detection and correction:** explanation capabilities help identify errors or mistakes made by the AI system. Users can understand why certain outputs may be incorrect or undesirable, allowing for improvements and corrections to be made.
- **Intellectual property and ownership:** Explainability can help establish ownership and intellectual property rights in AI-generated works. By understanding the creative process behind AI-generated content, individuals and organizations can assert ownership and defend their rights.

2.2 Protecting intellectual property and sensitive information

AI governance can help to protect corporate intellectual property by ensuring that it is properly identified and managed during the information processing lifecycle.

- **Protecting sensitive information:** AI models are often trained on sensitive data, such as customer data and financial data. By protecting the confidentiality of this data, organizations can avoid high-impact risks, such as data breaches, reputational damage, and regulatory fines.
- **To comply with regulations:** many regulations require organizations to protect intellectual property (IP) and sensitive information about their customers and employees. For example, the General Data Protection Regulation (GDPR) requires organizations to implement appropriate security measures to protect personal data.

2.3 Combating hallucination

Generative AI models can be used to create content that is sometimes indistinguishable from real content (hallucinate), which can lead to people being misled or deceived. When this happens, trust in AI and the institution that served the content can be undermined. Combating hallucination is hence

Table 1: Controls for the three key generative AI governance outcomes

KEY OUTCOMES OF OUR GENERATIVE AI GOVERNANCE MODEL		
BEING ABLE TO EXPLAIN THE RESULTS THAT AI PROVIDED	PROTECTING INTELLECTUAL PROPERTY AND SENSITIVE INFORMATION	COMBATING HALLUCINATION
Corresponding controls that are used to determine whether outcomes are being achieved		
1. There is a clear traceable connection between input (provided to AI as context) and the result returned by AI. 2. The AI system employs explainable AI techniques to provide interpretable explanations for its decisions. 3. The AI model is validated using specific transparency metrics to ensure its decision-making process is transparent. 4. The explanations provided by the AI system are audited by third-party experts to verify their accuracy. 5. Multiple AI models are employed to provide insights into the decision-making process, factors they consider, and the explanations they provide for their outputs. 6. Regular fairness/bias testing cycles are conducted. 7. Fairness-aware algorithms are employed during model training.	1. Any data sent to generative AI models is thoroughly vetted to ensure that it does not contain sensitive or proprietary information. Specifically, data minimization techniques are employed (only provide the LLM with the minimum amount of sensitive corporate material necessary for its intended purpose). 2. Data anonymization and redaction techniques are employed to remove any identifying information from the input data. 3. The generative AI model and information processing pipeline are deployed in an environment with restricted access, preventing unauthorized access to sensitive information. 4. Data usage audits are conducted regularly to verify compliance with intellectual property protection policies. 5. A data inventory is maintained. 6. Regular privacy audits to test for compliance are conducted.	1. Guardrails are applied to the generative AI models to prevent them from providing information outside set boundaries. 2. Generative AI models are tested on a diverse set of inputs, including edge cases and outliers, to verify that they do not generate unrealistic or nonsensical outputs. 3. An ensemble of generative AI models is used to cross-validate outputs and reduce the risk of hallucination. 4. Adversarial testing is performed to assess the model's resilience against potential hallucinatory inputs. 5. Generative AI models are continuously monitored in production to detect any potential cases of hallucination. 6. A real-time alerting mechanism is in place to notify responsible personnel if the AI model breaches guardrails. 7. In the event of guardrail breaches, a well-defined incident response plan is activated to investigate, rectify, and prevent similar incidents in the future.

Table 2: Application of the controls over the knowledge lifecycle

	PRE-PROCESSING	IN-PROCESSING	POST-PROCESSING
	Pertains to the first analysis of raw information that will ultimately be used to finetune generative AI models. The approach involves annotating, or marking up, raw data that will facilitate tracing output from Gen-AI.	Pertains to filtering and redacting the augmented raw input information towards ensuring that generative AI models receive the smallest amount of information required to perform its tasks.	Pertains to tasks that need to be executed as part of testing and monitoring AI-models. This also includes proactive steps that can be taken to ensure generative AI behaves within set parameters and boundaries.
Controls related to being able to explain the results that AI provided			
Controls related to protecting intellectual property and sensitive information			
Controls related to combating hallucination			

a critical governance objective and companies need to ensure the accuracy and reliability of Gen-AI models and work to build trust in AI. If end users cannot trust that generative AI models are producing accurate and reliable results, they are less likely to use them. This could hinder the adoption of the technology and realizing its potential benefits.

3. THE CONTROL FRAMEWORK

Defining a set of controls lies at the heart of any pragmatic governance model and represents the first step in building out our model. The controls define the mechanism by which one can comprehensively measure whether any given outcome is being achieved. A set of controls that can help achieve the three key outcomes described above is presented in Table 1.

It should be noted that corporate proprietary information and other sensitive data that are used to provide generative AI models with context towards assisting with specialized use cases generally go through a "knowledge lifecycle" that comprises of three steps: knowledge preparation (pre-processing), informing generative AI (in-processing), and knowledge consumption (post-processing). The controls that we defined in Table 1 should logically be applied during specific points across this lifecycle. Table 2 demonstrates this.

Applying all the controls listed in Table 1 is not a trivial task. To fully implement the generative AI governance model, processes need to be built around these controls so that they can execute and be reported against. To practically apply these controls, we strongly recommend maximizing automation around the supporting processes.

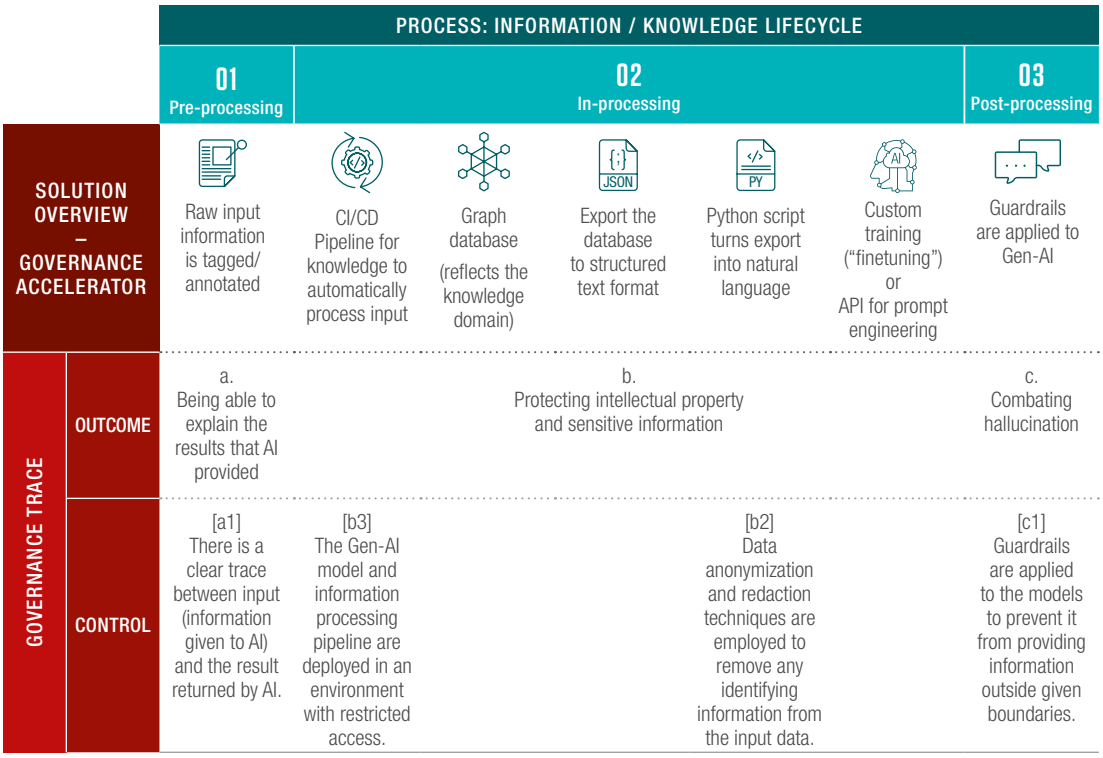
It should be mentioned, however, that while automation does not present a silver bullet towards AI-governance, it can significantly impact the cost and time footprint of executing many of the controls.

As part of our internal research, we have made significant advances in automating the execution of some of the control points listed in Table 1. These will be explored in the next section. We will also provide an overview of some of the work that other institutions are doing against some of the other control points we mentioned.

4. APPLIED AI-GOVERNANCE

Figure 2 provides an overview of a solution we have designed and built around the three "knowledge lifecycle" components we described earlier. The solution, which comprises of a collection of "control automation accelerators", traces to both the generative AI governance outcomes we defined as well as some of the control points that were listed in Table 1:

Figure 2: Generative AI governance automation solution overview



Our solution can be divided into three sub-solutions that directly correlate with the three key outcomes we described earlier:

4.1 Being able to explain the results that AI provides

Key to explaining AI's responses is to create a trace between generative AI's output and the material that was provided to it as input through finetuning.

To achieve this goal, we apply automated content markup and classification, through a couple of steps:

- We break the input knowledge that will be fed to AI into smaller fragments, such as pages or paragraphs (The raw input data may be in the form of large text-based documents, such as large PDF files, and be federated across multiple repositories).
- Metadata, in the form of keywords, are extracted from the fragments using RAKE (rapid automatic keyword extraction)[4] and synonyms of these keywords are obtained via consulting generative AI. This metadata, together with contextual information about the location (page number, paragraph number, information repository link, etc.) of the knowledge fragment is added to the input that is provided to AI.

During information retrieval:

- We dynamically match the user's query with the metadata we extracted and, towards better system performance, we only share content where we have a good match with generative AI.
- As part of this, we include the contextual information regarding where in the document the fragment comes from (page, paragraph, etc.).
- When AI responds to the user's query, it references the source by using a template to format generative AI's response in a way that includes the page and paragraph number.

[4] https://tinyurl.com/3xezt826

As a result, we can provide a detailed trace between the response and the specific information AI used to generate the response, which significantly facilitates explainability.

4.2 Protecting intellectual property and sensitive information

This outcome can be readily attained by using end-to-end automation in processing the information from its genesis point to where it is handed to AI as part of custom training or as prompt input. The key principle we apply here is that no human hands should touch the data.

To achieve this, we defined a solution that operates across the in-processing sub-process of the "information lifecycle process" (Figure 2):

- Harvests knowledge/information the moment it gets published into a version control system (such as Git or Subversion) through using a CI/CD (continuous integration/continuous delivery) pipeline for assets solution. The knowledge base that is harvested can be of any type – images, videos, audio, 3D models, data files, etc.
- Automation ensures that this information is processed, and a knowledge graph (which models/reflects the underlying knowledge domain) is updated accordingly.
- The knowledge graph is periodically exported, and the export file is converted to natural language through using a Python script, and from there it is injected automatically into a large language model (LLM) by using the LLM's "application programming interface" (API).
- Checks and balances along the processing pipeline ensure that what is sent to the generative AI model is appropriately filtered, redacted, or anonymized.

Because the entire process is automated, and access to the data in any stage of processing is highly restricted, corporate intellectual capital and other sensitive information is much better protected.

4.3 COMBATING HALLUCINATION

Hallucinations can be caused by a number of factors, such as the quality of the training data, the complexity of the model, and the way in which the model is used. To combat hallucination, we apply concepts we introduced earlier:

- **Use of external knowledge:** by incorporating contextual and external knowledge into the model, the likelihood of hallucination is reduced through providing AI with a more accurate representation of the world, within the context of the specific use case.
- **Data augmentation:** this technique involves transforming training data in various ways to expose the model to a wider range of patterns. By doing so, the model becomes more robust and less prone to hallucinating. In our practical example, we accomplished this by adding synonyms of key concepts that are addressed in source knowledge to the metadata that is used for prompting AI.

Additionally, we constrain AI through smart prompting (we script additional instructions and add it to the end user's input) to only employ the source knowledge we provided for constructing its responses. To do this, we take control of the entire user experience lifecycle and supplement user queries with these additional instructions in the background, i.e., explicit instructions to prevent AI from generating unrealistic or nonsensical responses. Finally, we set confidence thresholds for generated outputs. If the model's confidence falls below a certain threshold, the output can be flagged for further review or discarded to avoid potential hallucination.

5. A BRIEF OVERVIEW OF WHAT OTHERS ARE DOING

Several other institutions are conducting research in AI governance. We try and connect some of these endeavors to a subset of the controls we defined in Table 2.

5.1 Being able to explain the results that AI provides

5.1.1 CONTROL #2

The AI system employs explainable AI techniques to provide interpretable explanations for its decisions.

- Google AI has developed a number of explainable AI techniques, including LIME and SHAP (Google Colab).[5] These techniques are used in a variety of Google products, such as Google Search and Google Photos.
- Microsoft Research has also developed a number of explainable AI techniques (Explainability – Microsoft Research). These techniques are used in Microsoft products, such as Microsoft Azure Machine Learning and Microsoft Power BI.

[5] https://tinyurl.com/3ywd2cdw

- IBM Research develops and applies explainable AI techniques to a variety of problems (Explainable AI | IBM Research),[7] such as fraud detection and healthcare decision making.
- Amazon Web Services offers a number of explainable AI services, such as Amazon SageMaker Explainable AI.[8]

5.1.2 CONTROL #3

The AI model is validated using specific transparency metrics to ensure its decision-making process is transparent.

To ensure that generative AI models are used transparently, it is important to define and then validate them using transparency metrics. Some of the key terms involved in LLM transparency include:

- **Perplexity:** a measure of how well an LLM can predict the next word in a sequence.
- **Coherence:** a measure of how well the LLM's outputs make sense semantically.
- **Context appropriateness:** a measure of how well the LLM's outputs are relevant to the given context.

According to AIMultiple,[9] one of the key steps that organizations can take to validate LLMs for transparency is to use multiple evaluation metrics. Instead of relying solely on perplexity, for example, incorporate various evaluation metrics that capture different aspects of the LLM's performance, such as the ones we listed above. Moreover, it is important to implement transparency by design. One approach is using the "community transformer" design, which is a type of LLM architecture that is designed to offer a higher level of transparency than traditional LLM architectures. This design specifically allows users to see how the LLM is attending to different parts of the input sequence and how it is making its predictions.

5.1.3 CONTROL #7

Fairness-aware algorithms are employed during model training. Cornell University reported the following regarding employing fairness-aware algorithms[10]:

Familiarize yourself with different fairness definitions and metrics to identify the most suitable ones for your specific Gen-AI application.

Some "fairness definitions" include (but are not limited to) "demographic parity", the proportion of individuals from different protected groups (e.g., gender, race, ethnicity) who receive a favorable outcome should be equal; "individual fairness", which states that individuals who are similar in all relevant respects should receive similar outcomes, regardless of their protected group membership; and "counterfactual fairness", which states that individuals should receive the same outcome that they would have received if their protected group membership had been different.

Some "fairness metrics" include (but are not limited to) "discrimination ratio", which is calculated by dividing the proportion of individuals from a protected group who receive a favorable outcome by the proportion of individuals from a non-protected group who receive a favorable outcome, and "fairness-aware accuracy", which is calculated by taking the weighted average of the accuracy for each protected group, where the weights are determined by the size of each protected group.

Fairness metrics should be applied across different moments in the information lifecycle, including:

- **Pre-processing:** pre-processing techniques must be applied before data is fed to generative AI. This can include re-sampling, re-weighting, or transforming the data to ensure a more balanced representation of different groups.
- **In-processing:** incorporate fairness-aware optimization techniques during the LLM training process. These techniques can help balance the trade-off between model accuracy and fairness by adjusting the model's parameters or loss function.
- **Post-processing:** refers to post-processing techniques used to adjust the model's outputs to ensure fairness. This can include thresholding or calibration methods to achieve desired fairness metrics.

It is important to compare different fairness-aware algorithms and techniques to identify the most effective approach for your specific use case.

[7] https://tinyurl.com/3tm7kakp
[8] https://tinyurl.com/ywf64vt3
[9] https://tinyurl.com/mwphnjfn
[10] https://tinyurl.com/bdd5h5wn

5.2 Protecting intellectual property and sensitive information

5.2.1 CONTROL #5

A data inventory is maintained, which identifies, collects, and organizes personal data in systems, tracks data sources, and helps map how an organization's data assets are stored and shared. Although the concept of a data inventory is not new, it has gained prominence in recent years due to regulations like GDPR and CCPA (California Consumer Privacy Act), which require companies to have greater control over their data and to help organizations identify sensitive data.

According to RedClover Advisors,[11] a "data inventory" solution is predicated around data collection, usage, storage, and sharing practices; types of data collected; who data has been collected from; whether the data falls into any sensitive categories; and consent requirements.

Within the context of Gen-AI, some of the challenges in creating an efficient data inventory for "large language models" (LLMs) include:

- **Complexity of datasets:** LLMs require large volumes of data for training, which can make organizing and managing this data challenging.
- **Timeliness of information:** LLMs may not have updated information, as their knowledge is based on the training data available at the time of training.
- **Data source integration:** injecting knowledge into LLMs from various sources, such as external structured databases or company-specific APIs, can be challenging.
- **Data fragmentation and silos:** the existence of data silos and fragmentation of information across different platforms and systems can hinder the creation of a comprehensive and efficient data inventory.

To overcome these challenges and facilitate the creation of a data inventory, the following actions can be followed: implement a data warehouse or data lake to store all data used for training AI in a centralized location, which will make it easier to create and maintain a comprehensive and efficient data inventory; use a data management platform, such as Apache Hive, to help organize and manage large volumes of data; use a data pipeline to automate the data lifecycle, such as the CI/CD pipeline for assets solution we described earlier, together with version control to track changes to the data inventory; and use a data integration platform, such as Apache Nifi, to connect data silos and fragmentated information across different platforms and systems.

5.3 Combating hallucination

5.3.1 CONTROL #3

An ensemble of generative AI models is used to cross-validate outputs and reduce the risk of hallucination.

Robust Intelligence presented an approach for using an ensemble of generative AI models to reduce the risk of hallucination through:[12]

- Choosing a variety of generative AI models with different architectures, training data, or hyperparameters, diversity in their predictions can be ensured to reduce the likelihood of all the models hallucinating in the same way.
- Combining model outputs by using techniques such as voting (for classification tasks) or averaging (for regression tasks).
- Evaluating ensemble performance by using metrics relevant to the specific application/use case. Comparing the ensemble's performance to that of individual models is also important for ensuring that the ensemble is providing improved results.

5.3.2 CONTROL #6

A real-time alerting mechanism is in place to notify responsible personnel if the AI model breaches guardrails. According to Tata Consulting Services, a real-time alerting mechanism that notifies responsible personnel if the LLM breaches guardrails can be implemented using the following steps:[13]

- Establish clear guardrails for generative AI, i.e., a set of programmable constraints and rules that monitor and dictate user interactions with the model, ensuring it operates within defined boundaries and adheres to specific rules or principles. Examples of such guardrails include: "accuracy guardrail", which ensures that the AI model is performing as expected and is meeting its accuracy goals (should the model's accuracy fall below the set value an alert will be sent out); and "bias guardrail", which ensures

[11] https://tinyurl.com/mr32xvb2
[12] https://tinyurl.com/r2m2uyrt
[13] https://tinyurl.com/4b4e5hmk

that the AI model is not biased against any particular group or individual (the fairness metrics we referenced earlier can be used to define thresholds, which, if violated, will trigger the alerting mechanism).

- Continuously monitor the LLM's performance and outputs in real time, checking for any breaches of the established guardrails.
- Develop an alerting system that triggers notifications to responsible personnel when a breach of guardrails is detected. Email, SMS, and Slack messages are examples of potential alert carriers.

6. EMERGING ROLES AND EXECUTIVE PARTICIPATION

To support new processes that need to be built around the controls that we defined in Table 1, new roles will need to emerge. Here are a few examples:

- **AI Governance Lead:** this role will oversee the implementation and execution of the generative AI governance model and control set. The Lead will need to have a deep understanding of AI technology and the risks and challenges associated with its use in the financial services industry.
- **AI Risk Manager:** this role will be responsible for identifying and assessing the risks associated with the use of generative AI, and thus have a strong understanding of general risk management principles and practices.
- **AI Compliance Officer:** this role will be responsible for ensuring that the use of generative AI complies with all applicable laws and regulations, and needs to have a strong understanding of the legal and regulatory landscape for AI in the financial services industry.
- **AI Ethics Officer:** this role will be responsible for ensuring that the use of generative AI is ethical and responsible.
- **AI Technical Architect:** this role will be responsible for designing and implementing the technical infrastructure to support the generative AI governance model, and needs to have a deep understanding of AI technology stack and the associated infrastructure requirements.

Moreover, the implications of generative AI governance for CIOs, CTOs, CFOs, and business leaders are also significant. CIOs will need to ensure that the IT infrastructure is in place to support the generative AI governance model. This includes providing the necessary computing resources, data storage, and security controls. CIOs will also need to work with other stakeholders to develop and implement policies and procedures for the responsible use of AI.

CTOs will need to work with the AI Governance Lead to ensure that the generative AI governance model is aligned with the overall IT strategy. CTOs are ultimately responsible for implementing and maintaining the generative AI governance model as well as developing and deploying the necessary tools and technologies.

CFOs will need to budget for the costs of implementing and maintaining the generative AI governance model. This includes the costs of new roles, as well as the costs of new tools and technologies.

Business leaders need to ensure that the AI governance model is effective in meeting the needs of the organization. This includes understanding the importance of generative AI governance and being comfortable that AI solutions are being used in a way that aligns with the financial institution's values and principles towards building trust with customers, employees, and regulators; being involved in the development and implementation of the generative AI governance model to ensure that it is aligned with the organization's overall business strategy; and helping with monitoring and evaluating the generative AI governance model on an ongoing basis, since it needs to be adapted to changes in the regulatory landscape.

7. CONCLUSION

Generative AI presents additional challenges in the domain of AI governance, particularly around key outcomes such as transparency, protection of sensitive information, and combating hallucination. Defining a lean set of controls that trace to the outcomes and building supporting processes around these controls are at the heart of establishing a pragmatic governance model.

In our research, we were successful in partially achieving the desired outcomes by applying a combination of control automation in the information processing lifecycle, together with techniques to better contain generative AI within a clear set of boundaries to combat hallucination. We previously reported on this in an earlier article.[14]

Moreover, the need for AI-related governance is well recognized in the industry and many institutions have provided solutions around some of the controls we discussed. The solutions referenced in this paper together with our own governance accelerators collectively form an excellent primer for establishing a robust generative AI governance practice within an enterprise.

In closing, some final points key points about generative AI governance are that:

- Much of it is new and complex.
- It can radically change ways of working and how reliability is assessed.
- It involves many and very disparate stakeholders.
- It goes to the heart of key processes (such as client interactions, delivery at quality).
- It is not a one-time event. It is an ongoing process that needs to be adapted continuously to changes in AI technology and the regulatory landscape.

Hence, it is important to pay attention to testing the governance model as one develops it, much in the same way that for a project one needs to test the governance and delivery methodology.

Towards this end, we have the following recommendations when establishing a generative AI governance model within a financial institution:

- Not all Gen-AI use cases will require the same level of governance and control. Hence, we recommend defining a "minimum viable governance" (MVG) model on a case-by-case basis.
- Define and implement the MVG when the use case is in pilot phase already. This is because retroactively applying a robust governance structure is likely to result in significant churn in the core solution.

[14] https://tinyurl.com/27kp9d5e

ABOUT CAPCO

Capco, a Wipro company, is a global technology and management consultancy focused in the financial services industry. Capco operates at the intersection of business and technology by combining innovative thinking with unrivalled industry knowledge to fast-track digital initiatives for banking and payments, capital markets, wealth and asset management, insurance, and the energy sector. Capco's cutting-edge ingenuity is brought to life through its award-winning Be Yourself At Work culture and diverse talent.

To learn more, visit www.capco.com or follow us on Facebook, YouTube, LinkedIn and Instagram.

WORLDWIDE OFFICES

APAC
Bangalore – Electronic City
Bangalore – Sarjapur Road
Bangkok
Chennai
Dubai
Gurgaon
Hong Kong
Hyderabad
Kuala Lumpur
Mumbai
Pune
Singapore

EUROPE
Berlin
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
London
Milan
Munich
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA
Charlotte
Chicago
Dallas
Hartford
Houston
New York
Orlando
Toronto
Washington, DC

SOUTH AMERICA
Alphaville
São Paulo

THE COVER IMAGE WAS CREATED USING JASPER AI, AN AI ART GENERATOR

FSC
www.fsc.org
MIX
Paper from responsible sources
FSC® C004309

WWW.CAPCO.COM

CAPCO 25
a wipro company