ADI

Whitepaper

Navigating GenAI Ethics: A review of its regulatory landscape and a proposed governance framework for all stages of the GenAI development lifecycle

Gabriel Isaac L. Ramolete, Joshua B. Ramos, Adrienne Heinrich, Jesica M. Aaron, Romina Angeliz H. Marcaida, David R. Hardoon

Aboitiz Data Innovation (ADI)

About Aboitiz Data Innovation (ADI)

Aboitiz Data Innovation (ADI) is an AI solutions company reshaping industries with data science and AI. A startup with 150 years of legacy. With a proven track record across sectors like banking, financial services, power, and industrials, we deliver custom models, and GenAI platforms that drive outcome and impact.

Headquartered in Singapore, we boast a team of 100+ data scientists and AI algorithm engineers spread across 14+ countries. Built on our roots in Philippines' Aboitiz Group, our extensive capabilities and domain expertise uniquely positions us to serve our customers with unparalleled effectiveness and innovation.

adi.tech

Executive Summary

This paper presents a comprehensive review of current efforts to define GenAI regulatory and ethical standards, proposing a GenAI-specific framework for a wide range of stakeholders to navigate the ethical considerations inherent to GenAI. By establishing best practices through the lens of a typical GenAI development lifecycle, the proposed framework aims to foster a future where GenAI empowers humanity, regardless of industry, scale, or context.

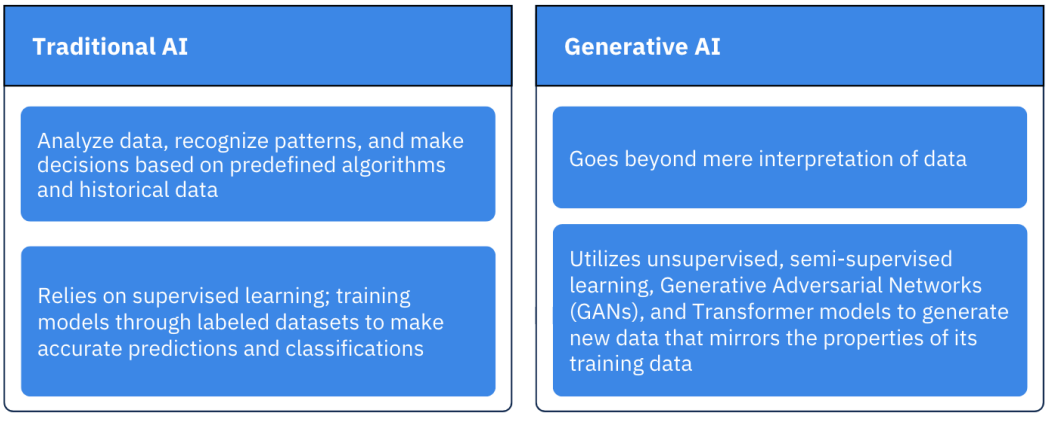
The discussion consists of case studies, frameworks and analysis outlined as follows:

- Discussion of definition & scope, importance of ethical governance in GenAI;
- Review & call to action;
- Proposed framework;
- Considerations when implementing the GenAI governance framework;
- Future directions and innovation in GenAI; and
- Conclusion

Introduction

Definition and Scope of GenAI

Distinguishing Traditional AI from Gen AI



The key distinction lies in GenAI's creative capacity, setting it apart from the more analytical and predictive nature of traditional AI systems. As Gen AI delves into a deeper analysis on

the creative capacity, it inhibits more sensitive areas in terms of risks and information; thus, calling for a collaborative and responsible approach to GenAI systems.

Review and Call to Action

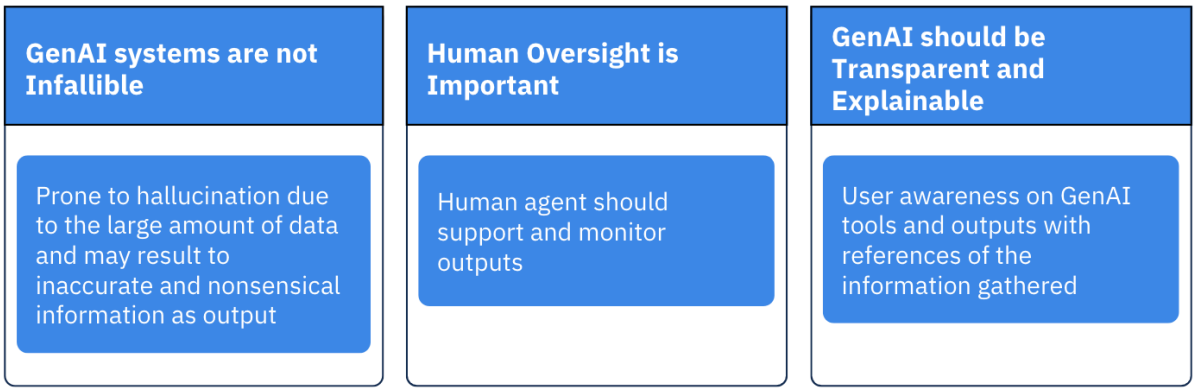
Definitions and Terms

What is “Ethics” in the context of this framework? → A comprehensive set of values and principles that guide the development, deployment, and use of GenAI systems in a way that benefits society, minimizes potential harm, and aligns with our fundamental moral principles.

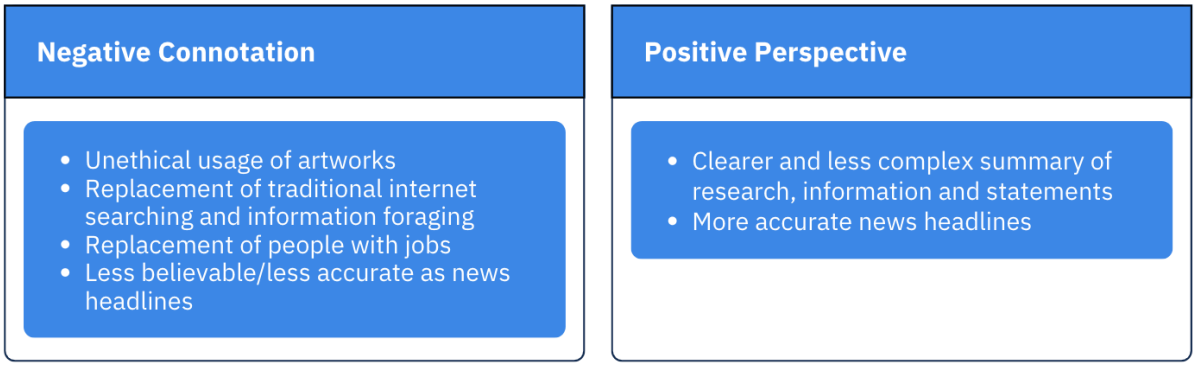
Differentiating “ethical”, “harmful”, and “unethical”:

- Ethical
    - Alignment with our eleven (11) core principles
    - Societal benefit
    - Risk minimization
    - Ex. Generating high-resolution images of existing satellite images to identify deforestation patterns
- Harmful
    - Negative consequences
    - Lack of safeguard
    - Neglect of ethical considerations
    - Ex. Content filtering and summarization for posts in social media platform, but algorithm removes legitimate content and allows harmful misinformation to spread
- Unethical
    - Intentional violation of principles
    - Disregard for social impact
    - Ex. GenAI chatbot that manipulates customers into making unnecessary purchases

Lessons Learned from Ethical Failures



Public Perception and Trust



Presenting GenAI Frameworks Across the Globe

GenAI holds potential for innovation across various sectors, but its development necessitates various ethical, legal, and cultural considerations. To address these concerns, various researchers and organizations have proposed and established GenAI frameworks, from universal to industry-specific use. This section explores prominent frameworks from across the globe to understand their diverse approaches to governing GenAI creation.

Frameworks from Countries and Organizations

Singapore's Model AI Gov Framework

Internal governance structures and measures
- Multidisciplinary body to oversee, develop guidelines/standards in designing & implementing AI systems responsibly

Human involvement in AI-augmented decision-making
- Conduct of risk impact assessments & defining human involvement

Operations Management
- Risk assessment before data collection or modelling

Stakeholder interaction and communication
- Ensure stakeholder awareness throughout the lifecycle

AI Risk Management Framework (NIST)

Govern
- Establish a culture of risk management through the lifecycle – identify & manage risks and define roles & responsibilities

Map
- Understand the AI system through characterizing its intended use and the environment it operates in

Measure
- Assess the likelihood and impact of potential risks identified

Manage
- Develop plans to address identified risks, like implementing controls to mitigate risks, monitoring system performance, and continuously improving risk management processes

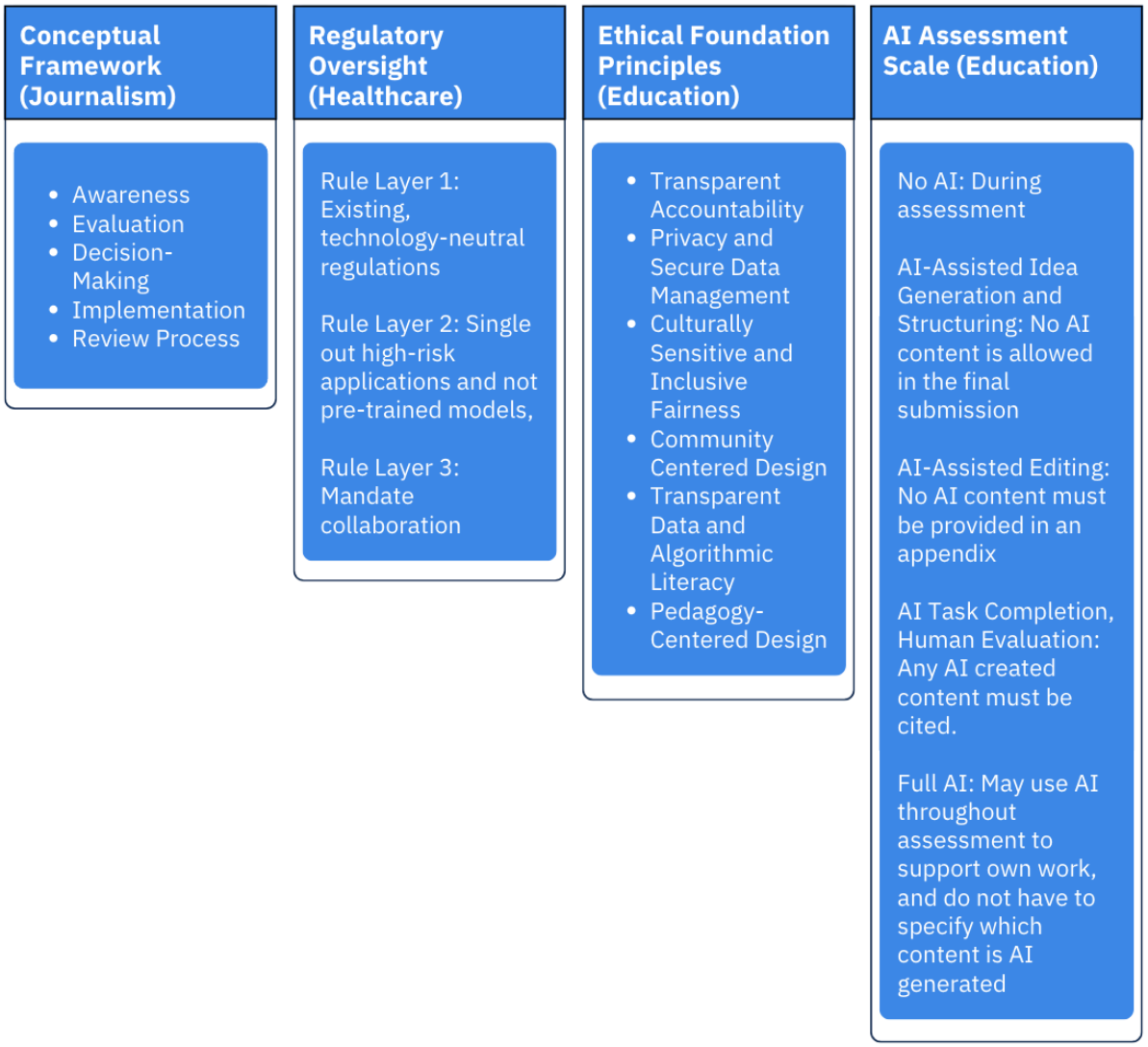
Risk Categories (EU AI Act)

Unacceptable Risk (Prohibited)
- Systems deemed too risky, like social scoring for government control or subliminal manipulation and deception, are banned.

High-Risk
- Applications with significant potential for harm, such as recruitment tools and biometric ID systems, face stricter regulations to ensure fairness, transparency, and human oversight. Generative AI tools are likely to fall under this bracket.

Low-Risk
- Applications with minimal risk, such as spam filters and AI-powered games, seek minimal regulation.

Frameworks from Research Communities



What is still needed?

Most frameworks related to Generative AI tools and implementations are on general purpose and traditional AI systems. While these have significant merit, there is still a need for clarity on whether Generative AI systems and tools should be held under the same guidelines and regulations, or if additional and more intricate measures should be suggested.

It is our belief as Aboitiz Data Innovation (ADI) that there is still a need for separate terminologies, objectives, and action points for generative AI-based systems, due to its unique nature of generating content rather than predicting. We list down below why our framework is essential for the responsible development and deployment of GenAI:

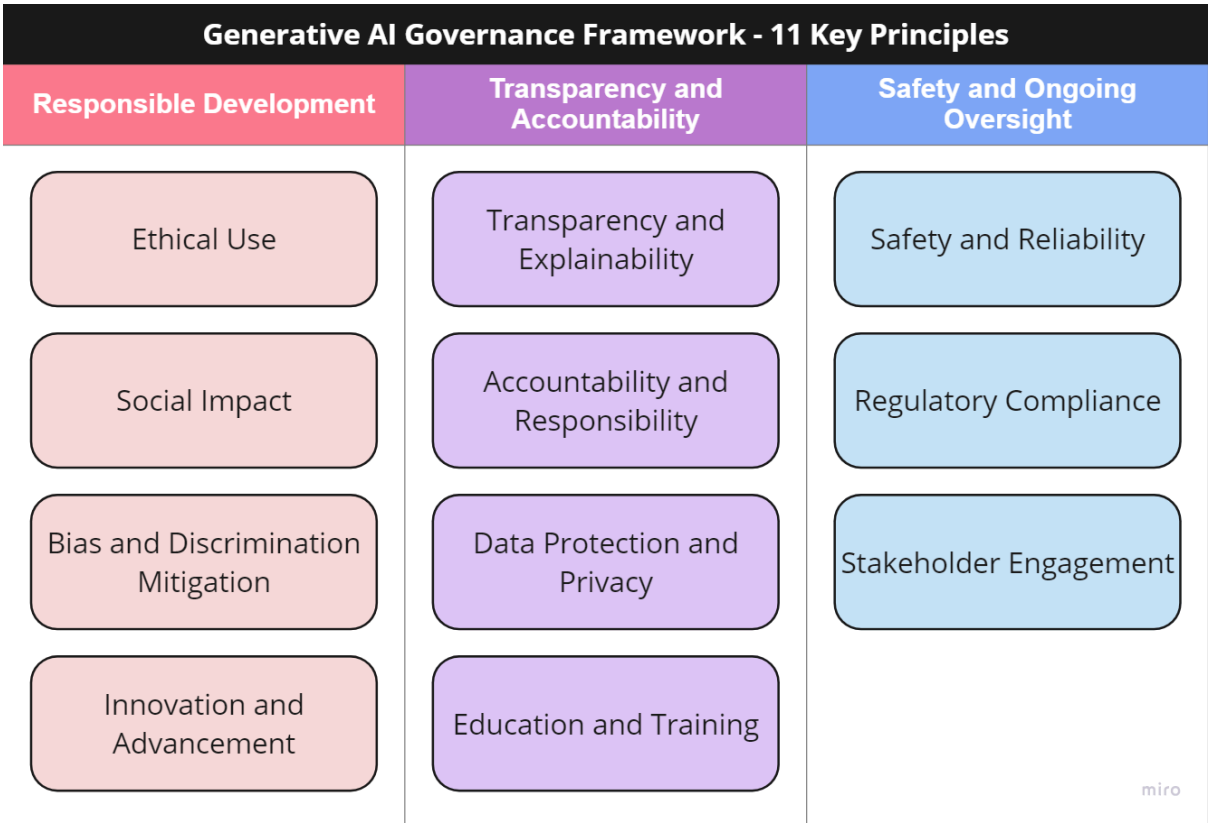
1. Lifecycle-based focus – our framework delves into each stage of the GenAI lifecycle – from problem definition to ongoing monitoring – highlighting the critical principles that require specific focus at each juncture.
2. Mitigating risks across industries – incorporate beneficial and harmful examples across various industry use cases, allowing stakeholders to identify potential risks and opportunities specific to their domain, fostering a nuanced and stricter approach compared to general-purpose AI
3. Framework vs. No Framework: A clear distinction in outcomes
    a. GenAI with Framework: Despite increased development time, there will be reduced bias, improved transparency, responsible data usage, ethical alignment, positive societal impact
    b. GenAI w/o Framework: Even with potentially faster initial development timeline, there will likely be more unintended bias, safety vulnerabilities, privacy violations, unethical applications, and negative societal consequences

The Proposed Framework

The principles to be presented are summarized based on best practices on deployment and implementation of GenAI systems.

Governance Principles

To attain a balanced and comprehensive approach to managing the risks and benefits associated with GenAI, ensuring its development and use are aligned with societal values and ethical standards. The 11 Key Principles are grouped into three (3) categories.

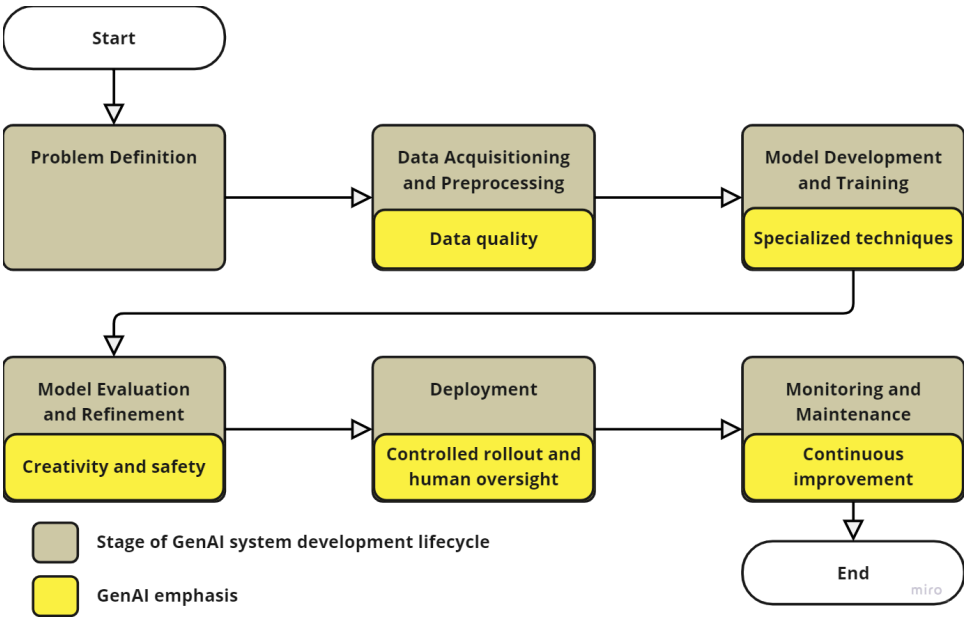


11 Key Principles of the Generative AI Governance Framework

The GenAI system development lifecycle

The GenAI system development lifecycle places a stronger emphasis on data quality, utilizing specialized model architecture, and prioritizing creative output and safety considerations during evaluation and deployment.

The lifecycle of a Generative AI system may share many similarities with a traditional or general-purpose AI system, but the key differences lie in data quality and additional human oversight, due to the scalability and added potential risks an GenAI system poses.



GenAI System Development Lifecycle

We can cross-apply principles to each lifecycle stage to apply holistic governance:

Stage	Principles	Considerations
Problem Definition	• Ethical Use • Social Impact • Stakeholder Engagement	• Clearly define purpose and intended use of GenAI system • Consider potential social implications and benefits with risks • Identify relevant stakeholders and involve them in early discussions
Data Acquisition and Preprocessing	• Bias and Discrimination Mitigation • Data Protection and Privacy • Transparency and Explainability	• Source data responsibly and ethically, considering potential biases • Implement robust data anonymization and privacy protection measures • Document data sources and preprocessing methods for transparency
Model Development and Training	• Bias and Discrimination Mitigation	• Choose training data and algorithms that minimize bias

	• Transparency and Explainability • Accountability and Responsibility	• Develop methods to explain model outputs and decision-making processes • Clearly define roles and responsibilities for model development and deployment
Model Evaluation and Refinement	• Bias and Discrimination Mitigation • Transparency and Explainability • Safety and Reliability	• Evaluate models for potential biases and fairness in outputs • Develop methods for assessing model explainability and interpretability • Test for potential safety risks and unintended consequences of model outputs
Deployment	• Ethical Use • Transparency and Explainability • Accountability and Responsibility • Regulatory Compliance	• Deploy the model in a way that aligns with its intended purpose and ethical principles • Ensure transparency regarding use of GenAI in deployed system • Clearly define lines of accountability for model performance and potential harm • Comply with all relevant regulations regarding GenAI deployment and data usage
Monitoring and Maintenance	• Bias and Discrimination Migration • Safety and Reliability • Transparency and Explainability • Accountability and Responsibility	• Continuously monitor model performance to detect and address bias drift • Monitor for safety issues and potential vulnerabilities in the deployed system • Maintain transparency regarding model updates • Be prepared to address any harm caused by the model and take corrective actions • Establish committees or approval boards where monitoring information is either reported or available to

The six-stage GenAI system development lifecycle and corresponding GenAI Governance principles

This framework integrates the identified principles into the six-stage Generative AI system development lifecycle. Other principles, such as Innovation and Advancement, Stakeholder Engagement, and Education and Training, may be relevant throughout the entire lifecycle, as we foster a culture of continuous learning and responsible innovation.

Ethics of GenAI with Several Industries through the lens of the GenAI Governance Framework

One of the sectors affected by GenAI development is the education sector. Many students use GenAI for the tasks and requirements to generate content. According to the Turnitin company, 22 million papers are suspected to be AI generated for the past year. However, some researchers argue that utilization of GenAI in school requirements is not necessarily unethical as long as the tools are used for valid and legitimate research and students are transparent to what tools they use. Maastricht University in Netherlands integrated GenAI in their academic settings by enhancing educational practices particularly problem-based learning (PBL) environments. With the data available in ChatGPT, Maastricht University fosters opportunities for students and educators to explore, critique, and refine AI-generated outputs within their learning processes.

In the manufacturing sector, GenAI helps streamline processes, optimize resource utilization and implement advanced quality control measures. Although most people's concern, especially in the manufacturing industry is replacement of workers, GenAI logically assists human working with automation, but human agents are still essential to business processes enabling workers to upskill to keep their jobs. By adhering to the guidelines mentioned in the table, manufacturers can harness the potential of GenAI for innovations in product design, predictive maintenance, and supply chain optimization, while also ensuring ethical and safe practices.

Numerous companies in the energy sector are prioritizing the development and deployment of GenAI as a strategic imperative (Velasco, 2024). This trend reflects a growing recognition of GenAI's transformative potential in optimizing operations, enhancing efficiency, and supporting sustainability goals across the industry. Example real-life use cases of GenAI in

this industry are load forecasting, power outage prediction and preventive maintenance. As these introduce ethical considerations, a suggested framework is also presented.

The banking center achieves great heights with the integration if GenAI increasing significant opportunities in enhancing customer interaction through advances chatbot functionalities, improves fraud detection capabilities, and automates complex tasks such as code development and regulatory reporting and credit score analysis. Notwithstanding these gains, integration of GenAI posits several risks and these may be mitigated through the suggested framework.

Considerations when Implementing the GenAI Governance Framework

The successful implementation of the Framework requires a collaborative approach – this section delves into strategies for fostering robust communication and collaboration among stakeholders across all stages of the GenAI development lifecycle.

Stakeholder Engagement

Identifying Key Stakeholders

In the typical GenAI system development lifecycle, there is a wide range of key stakeholders involved. Depending on the system and stage of development, some of these stakeholders may be more directly entangled than others. Depending on the specific GenAI system being developed, additional stakeholders might be involved, such as domain experts, investors, ethicists, public interest groups, or specific communities impacted by the technology.

Strategies for effective engagement

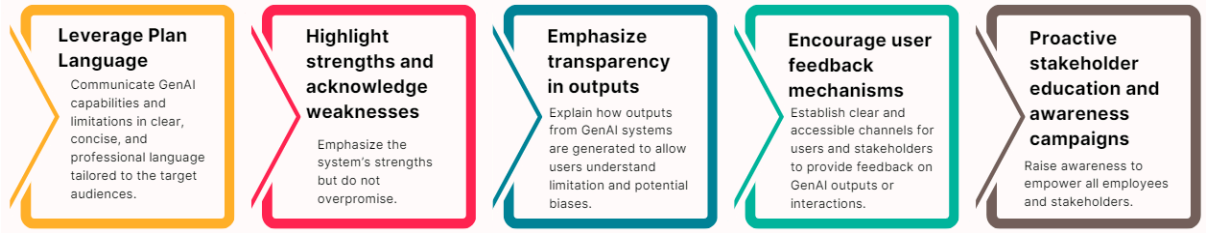
The successful and responsible development of a GenAI system necessitates a collaborative approach throughout the entire lifecycle. In Table 2, we detail the different responsibilities and strategies each stakeholder group can employ throughout the six stages of GenAI development.

Stage \ Stakeholder Group	Developers and Researchers	Businesses and Organizations	Regulatory Bodies and Policymakers	End Users and Consumers
Problem Definition	• Leverage technical expertise to identify potential applications of GenAI and define the specific problem the system aims to address • Researchers inform problem definition, ensuring chosen approach is technically feasible and ethically sound	• Provide strategic direction and establish desired functionalities for GenAI system	• Offer guidance on ethical considerations • Advise potential societal impacts of the proposed GenAI application	
Data Acquisition and Preprocessing	• Acquire data relevant to the problem definition • Prioritize data quality and address potential biases through careful selection and pre-processing techniques	• Provide legal expertise, advising on data privacy regulations and ensuring that data collection adheres to ethical and legal standards • Safeguard user privacy	• Offer best practices for data anonymization and responsible data handling • Contribute to ethical data management practices	
Model Development and Training	• Develop and train the GenAI model • Select algorithms and training data that minimize bias, mitigating potential	• Provide essential resources and oversight • Define how GenAI model fits into application or system		

	discriminatory outputs			
Model Evaluation and Refinement	• Evaluate and refine the model's performance in terms of accuracy, fairness, and potential biases	• Align expectations and outcomes with model's performance in terms of business goals	• Analyze model outputs for ethical concerns and recommend mitigation strategies • Recommend metrics and evaluation frameworks for assessing bias and fairness in GenAI models	
Deployment		• Oversee deployment of GenAI system, ensuring it aligns with intended purpose and it delivers the desired functionalities • Guarantee compliance with relevant regulations on AI deployment and data usage		• Receive clear information on use of GenAI within the system they interact with • Report issues and concerns to ensure ongoing improvement
Monitoring and Maintenance	• Monitor model's performance to potentially address bias drift and safety risks that may emerge over time	• Allocate resources for ongoing maintenance	• Review and adapt regulations, keeping pace with other technological advancements	

Transparent Communication Practices

Clear and honest communication about GenAI capabilities and limitations



By implementing these strategies, stakeholders can cultivate a culture of open and honest communication around GenAI capabilities and limitations. This transparency engenders trust with users, empowers them to leverage the technology responsibly, and paves the way for a future where GenAI serves as a powerful tool for societal good while mitigating potential risks.

Reporting and disclosure practices

Strategic communication is two-way, and it can be through surveys, town hall sessions, interviews, and design thinking workshops. Although these methods may build the foundation of the stakeholder management, building the strategy afterwards shall but put together and aligned by identified participants who has the role of the sharing the strategic plan moving forward. Towards implementation, it is important to proactively follow the stakeholder map and engage the appropriate stakeholders. Finally, monitoring stakeholder feedback is necessary to be able to adjust and align the approach in the GenAI development process.

Regulatory and Legal Considerations

This section explores the current state of GenAI-related regulations across various jurisdictions and proposes action points on how stakeholders can proactively address compliance challenges.

Compliance with Existing Laws and Regulations

1. European Union: The EU AI Act is a comprehensive framework that classifies AI systems into four tiers based on their risk levels: unacceptable, high, limited, and minimal risk. This classification considers the sensitivity of the data involved and the AI use case. Key prohibitions include the use of AI for manipulative, deceptive, or

subliminal techniques, exploitation of vulnerabilities, and biometric data categorization for discriminatory purposes.

2. United States: The US follows a decentralized approach with sector-specific regulations. Agencies like the Federal Trade Commission (FTC) and the National Highway Traffic Safety Administration (NHTSA) address consumer protection and autonomous vehicle safety, respectively. States such as California have additional regulations like the California Consumer Privacy Act (CCPA), which imposes strict data processing requirements.

3. China: China's AI strategy emphasizes data protection and risk management through frameworks such as the Cybersecurity Law and the New Generation AI Development Plan. The country's approach balances AI innovation with ethical practices to maintain its leadership in the AI sector.

4. Canada: Canada’s regulatory landscape includes the Pan-Canadian AI Strategy and the Canadian AI Ethics Council. The Personal Information Protection and Electronic Documents Act (PIPEDA) ensures stringent data protection and privacy rights in AI applications.

Organizations need a multifaceted approach to compliance, involving best practices, governance frameworks, sector-specific regulations, and territorial laws. Key actions include revisiting internal policies (e.g., privacy and data use policies), ensuring cross-functional collaboration within the organization, and consulting regulatory agency guidance. Engaging with policymakers is crucial for shaping balanced and effective AI regulations. Organizations can contribute by participating in government working groups, collaborating with experts, and showcasing responsible AI practices. This proactive engagement helps in developing new standards and guidelines that ensure the ethical and effective use of GenAI.

Proactively Shaping Future Regulations

Through proactive engagement with policymakers and stakeholders, organizations like ours, Aboitiz Data Innovation (ADI), can influence the creation of balanced and effective regulations that support the responsible development and deployment of generative AI technologies. ADI's involvement in government technical working groups contributes to the implementation of the Philippines National AI Strategy Roadmap, which focuses on digitalization, infrastructure, workforce development, regulation, and research and development. Additionally, ADI collaborates with AI and legal experts, think tanks, academia, and industry leaders to develop well-informed policy recommendations, presenting a unified front with broad support. By showcasing responsible AI practices and

participating in initiatives like the ASEAN Guide on AI Governance and Ethics, ADI demonstrates its commitment to ethical AI development and contributes to creating robust, interoperable standards that benefit society while managing risks.

Future Directions and Innovations in Generative AI Ethics

In this section, we investigate emerging areas of research and discussion that hold promise for mitigating ethical risks associated with AI, and how potential shifts in the regulatory landscape could affect the Framework and GenAI stakeholders.

Technological Advancements

Generative AI (GenAI) is on the cusp of a major leap forward. Improved deep learning techniques will enable models to generate highly complex, accurate, and realistic content. Advancements in VAEs and GANs promise even more diverse and impressive outputs. Smaller, more efficient language models will democratize AI, making it accessible to a broader audience. Open-source models will further fuel innovation.

GenAI is poised to move beyond siloed applications, embracing cross-domain creativity. We'll see models that seamlessly process and generate content across modalities like text, images, and music. This will unlock richer applications and drive innovation across industries. These models have the potential to revolutionize content creation, idea generation, and even collaborative creativity.

However, with this power comes peril. Increasingly autonomous GenAI systems raise ethical concerns, particularly regarding adversarial uses like undetectable deepfakes. The ability to generate personalized misinformation is especially troubling due to its potential to manipulate and exploit vulnerabilities. As GenAI evolves, we must find ways to harness its potential while mitigating these emerging risks.

Emerging Trends in GenAI Ethics

Data Responsibility and Privacy

- Safeguarding User Data: Importance of robust data governance and transparency in data collection, usage, and retention.
- Balancing Performance and Privacy: Need to balance AI model performance with privacy protection, considering ethical implications of using personal data.

Fairness, Transparency, and Inclusion

- Mitigating Bias: Designing AI models to avoid discriminatory outcomes and using techniques to identify and mitigate biases.
- Creating Fair and Representative Content: Ensuring generated content is free from stereotypes and biases, with a focus on transparency through Explainable AI (XAI).
- Inclusion and Accessibility: Ensuring AI benefits all users, including those with disabilities, by generating inclusive and culturally sensitive content.
- Ethical Metrics: Integrating fairness metrics like Disparate Impact and Equal Opportunity Difference into AI model design to evaluate ethical performance.
- Environmental Sustainability: Reducing AI's environmental impact through energy-efficient training, hardware optimizations, and responsible resource usage.

Accountability, Trust, and Regulatory Landscape

- Establishing Accountability Mechanisms: Responsibility for AI behavior through regular audits, documentation, and adherence to ethical guidelines, with XAI to build trust.
- Human-in-the-Loop: Ensuring human oversight in GenAI workflows for responsible decision-making in high-stakes applications.
- Transparent Communication: Importance of transparency and corrective actions when issues arise.
- Ownership and Misuse: Addressing questions of ownership, attribution, and potential misuse of realistic AI-generated content.
- Regulatory Influence: Anticipating evolving regulations and the need for user education and awareness about GenAI capabilities and limitations.

Key aspects of GenAI Ethics

Long-Term Ethical Considerations and Global Perspectives

GenAI presents a complex picture for human identity and society, varying across different cultures and regions and significantly impacting humans in both positive and negative ways:

- The Nature of Creativity and Originality: Will AI-generated works be considered true art, and if so, who deserves credit: the programmer, the AI itself, or a combination?
- The Future of Work and Automation: We need ongoing discussions about the future of work, including the need for reskilling and upskilling initiatives to ensure a smooth transition for individuals whose jobs are impacted.
- Ethical Considerations for Social Justice and Equality: How can we ensure that the benefits of GenAI are distributed fairly across different demographics and socioeconomic groups?
- The Impact on Human Values and Decision-Making: As GenAI becomes more integrated into our lives, how will it influence our values and decision-making processes?
- The Global Landscape of GenAI Ethics: The ethical considerations surrounding GenAI may differ based on societal values and existing regulations. International

collaboration on GenAI ethics is essential to ensure responsible development and deployment on a global scale.

- The Potential for Existential Risks: Some experts raise concerns about the long-term potential for highly advanced AI to pose existential risks to humanity.
- The Need for Continuous Discourse and Adaptation: The ethical landscape of GenAI is constantly evolving. It is essential to adapt ethical frameworks as technology advances and societal values change.

Conclusion

All in all, we conclude that our proposed GenAI Governance Framework provides a critical roadmap for responsible GenAI development and deployment. Its lifecycle-based approach, domain-specific considerations, and clear distinction between its benefits and those of a framework-less approach makes it an essential tool for fostering a responsible and inevitable future with GenAI.

Our proposed GenAI Governance Framework integrates eleven (11) key principles throughout the six (6) stages of the GenAI system development lifecycle. Considering these principles at each stage is crucial because it ensures ethical considerations are embedded throughout the entire process, from initial concept to ongoing monitoring. These principles can be grouped into three categories, providing a roadmap for navigating GenAI development and deployment:

- Responsible Development and Use: This category focuses on ensuring GenAI is developed and used ethically, considering societal impacts and potential biases. It encompasses principles such as:
    - Ethical Use
    - Social Impact
    - Bias and Discrimination Mitigation
    - Innovation and Advancement

- Transparency and Accountability: This category emphasizes the importance of clear communication and establishing lines of responsibility throughout the GenAI lifecycle. It includes principles like:
    - Transparency and Explainability

    - Accountability and Responsibility
    - Data Protection and Privacy
    - Education and Training

- Safety and Ongoing Oversight: This category focuses on ensuring the safety and reliability of GenAI systems, including addressing potential risks. It includes the principles:
    - Safety and Reliability
    - Regulatory Compliance
    - Stakeholder Engagement (for ongoing feedback and adaptation)

Navigating GenAI Ethics: A review of its regulatory landscape and a proposed governance framework for all stages of the GenAI development lifecycle

Gabriel Isaac L. Ramolete, Joshua B. Ramos, Adrienne Heinrich, Jesica M. Aaron, Romina Angeliz H. Marcaida, David R. Hardoon

Aboitiz Data Innovation (ADI)

Introduction

The field of Generative AI (GenAI) stands at a pivotal juncture. Its potential to revolutionize countless sectors, from scientific inquiry to artistic expression, is astounding – however the introduction of new opportunities equally raises the possibility of new risks. It is imperative proactively lay the proper governance foundations to ensure the ethical, unbiased, and socially responsible development and deployment of GenAI tools and systems.

This paper presents a comprehensive review of current efforts to define GenAI regulatory and ethical standards, proposing a GenAI-specific framework for a wide range of stakeholders to navigate the ethical considerations inherent to GenAI. By establishing best practices through the lens of a typical GenAI development lifecycle, the proposed framework aims to foster a future where GenAI empowers humanity, regardless of industry, scale, or context.

Definition and Scope of GenAI (Generative AI)

Generative AI (GenAI) is arguably one of the most renowned inventions in technology today.

According to IBM Research (Martineau, 2023), GenAI systems are advanced computer programs that can create realistic high-quality content by learning from large amounts of data. The content that they can generate ranges from text, images, sound, videos, or any combination of these four.

While traditional AI systems excel at analyzing data, recognizing patterns, and making decisions based on predefined algorithms and historical data, GenAI takes this capability a step further by creating novel content rather than merely interpreting existing data. Traditional AI relies heavily on supervised learning, where models are trained on labeled datasets to make accurate predictions or classifications. In contrast, GenAI utilizes techniques like unsupervised and semi-supervised learning, as well as advanced neural networks such as Generative Adversarial Networks (GANs) and Transformer models, to generate new data that mirrors the properties of its training data (Goodfellow, et al., 2014; Vaswani, et al., 2017). This ability to produce generated content makes GenAI particularly useful in creative fields such as art, music, and writing, as well as practical applications like data augmentation, design prototyping, and creating synthetic data for training other AI models (Brown, et al., 2020). Thus, the key distinction lies in GenAI's creative capacity, setting it apart from the more analytical and predictive nature of traditional AI systems.

Different GenAI algorithms have distinct components that enable them to produce novel content. One prominent type is Generative Adversarial Networks (GANs), which consist of two neural networks: a generator and a discriminator. The generator creates new data instances, while the discriminator evaluates them against real data, providing feedback that helps the generator improve (Goodfellow, et al., 2014). Another significant algorithm is the Transformer model, which relies on self-attention mechanisms to process input data in parallel, making it highly effective for generating coherent and contextually relevant text. The Transformer architecture underpins models like GPT-3, which uses layers of attention mechanisms to generate human-like text based on vast amounts of training data (Brown, et al., 2020; Vaswani, et al., 2017). Additionally, Variational Autoencoders (VAEs) are used in GenAI for generating data that is like the input data by learning a compressed representation of the data and then reconstructing it (Kingma & Welling, 2013). These diverse algorithms highlight the versatility of GenAI, enabling it to create high-quality, realistic content across various domains.

Importance of Ethical Governance in GenAI

The rapid increase in data and computing power has driven the development of data-centric technologies like AI. Organizations can leverage AI to offer innovative products and services, improve productivity, and enhance competitiveness, which can result in economic growth and an improved quality of life. However, like other emerging technologies, AI may bring new ethical, legal, and governance challenges. These challenges include the risk of amplified unintended discrimination that may result in unfair outcomes, and concerns about consumers being unaware of AI's role in making important or sensitive decisions about them (Personal Data Protection Commission Singapore, Infocomm Media Development Authority Singapore, 2020).

As generative AI tools become more popular, their use in the workplace is also increasing. Employees across various fields and levels are utilizing GenAI tools, making them relevant to almost every job. Consequently, organizations need to address the legal and social risks, benefits, and long-term implications of supporting and using generative AI. Ethical governance in GenAI is vital to ensure that (1) bias and discrimination is prevented, (2) there is trust and transparency, (3) there is accountability, (4) data protection and privacy is observed and practiced, (5) it supports long-term sustainability, and (6) compliant with laws and regulations.

Objectives

This paper serves as a call to collective action by introducing a collaborative and responsible approach to Generative AI systems. The objectives of this paper are as follows:

- Describe the current ethical and regulatory landscape of Generative AI through existing frameworks and case studies
- Propose a comprehensive Generative AI Governance framework
- Highlight considerations when implementing the framework with GenAI systems
- Encourage collaboration and public awareness for shaping a future with Generative AI

Review and Call to Action

Definitions and Terms

In the context of this framework, we refer to *"Ethics"* as a comprehensive set of values and principles that guide the development, deployment, and use of GenAI systems in a way that benefits society, minimizes potential harm, and aligns with our fundamental moral principles. By adhering to ethical principles, which we will be defined later in this paper, stakeholders involved in GenAI can ensure that this powerful technology is used for good and contributes to a more just and equitable world.

Below is a breakdown of how stakeholders can differentiate between "ethical", "harmful", and "unethical" in the context of GenAI:

Ethical:

- Alignment with Core Principles: An ethical GenAI system demonstrably adheres to the principles outlined in the framework. This encompasses aspects like responsible development, transparency and accountability, and ongoing oversight. This translates to proactive bias mitigation, robust data privacy protection, and clear communication regarding limitations and functionalities.

- Societal Benefit: An ethical GenAI system demonstrably delivers a positive social value. It could be harnessed for scientific discovery, personalized learning interventions, or advancements in healthcare delivery.
- Risk Minimization: Ethical GenAI development prioritizes proactive identification and mitigation of potential risks associated with the technology. This includes addressing bias, ensuring safety vulnerabilities are addressed, and preventing manipulation.

For example, A GenAI system can be made to generate higher-resolution images of existing satellite images to identify deforestation patterns. This information empowers environmental organizations to advocate for conservation efforts, exemplifying ethical use of GenAI that aligns with core principles and contributes to positive societal change.

Harmful:

- Negative Consequences: A GenAI system, even unintentionally, can be harmful if it exacerbates existing societal problems or creates new ones. These consequences could manifest as biases in hiring decisions, manipulation of online user behavior, or privacy violations.
- Lack of Safeguards: A harmful GenAI system might lack adequate safeguards to prevent potential misuse or unintended consequences. This could be due to a lack of bias mitigation techniques or insufficient safety testing during development.
- Neglect of Ethical Considerations: A harmful GenAI system might be developed without due consideration for ethical principles. This could involve prioritizing profit over ethical considerations or neglecting privacy concerns.

To illustrate, we give an example of a social media platform deploying a GenAI system for content filtering, but the algorithm inadvertently removes legitimate content while allowing harmful misinformation to spread. This can have negative consequences for public discourse and social harmony, demonstrating how unintended consequences can result in harmful outcomes.

Unethical:

- Intentional Violation of Principles: An unethical GenAI system is one that deliberately disregards ethical considerations for personal gain or malicious

purposes. This could involve manipulating user data for targeted advertising campaigns, creating deepfakes to spread misinformation, or developing autonomous weapons systems.

- Disregard for Social Impact: An unethical GenAI system prioritizes its purpose over potential negative societal impacts. This could be seen in a system designed to exploit user vulnerabilities or perpetuate social biases.

An unethical use case would be an advertising company that develops a GenAI system that predicts consumer behavior with the goal of manipulating them into making unnecessary purchases. This is a clear violation of ethical principles and prioritizes profit over consumer well-being, exemplifying a demonstrably unethical application of GenAI technology.

By leveraging the GenAI Governance Framework, stakeholders can navigate the complexities of this evolving technology and ensure its development and deployment are demonstrably ethical, beneficial, and minimize potential harm. This framework fosters a future where GenAI serves as a powerful tool for positive societal change.

Case Studies and Current Usage

This section highlights real-world examples of Generative AI usage, so that principles for responsible GenAI development are highlighted. We analyze the most practical applications of GenAI and highlight a series of use cases on how GenAI is being used in various industries.

To start, the most prominent of the GenAI model types are text-based models, or Large Language Models (LLMs). The most popular pioneer of LLMs is OpenAI's ChatGPT. Back in 2020, OpenAI successfully launched the base model of ChatGPT, the Generative Pre-trained Transformer 3 (GPT-3). Then, last November 2022, they officially released its conversational platform, the actual ChatGPT, which made use of the GPT-3 as its underlying technology (Chiu, 2024). A few months after its release, ChatGPT has become the fastest growing platform in terms of users, reaching 100 million users in just two months, beating both TikTok and Instagram, which reach those numbers in 9 months and 30 months (i.e., about 2 and a half years) respectively (Hu, 2023). With the success of ChatGPT, the other leading technology enterprises also followed its lead, releasing their very own LLMs (e.g., Google's Bard, Microsoft's Copilot, etc.).

The latest version of GPT today, the GPT-4, possesses the capability to process and output images (Clark & Vincent, 2023). In other words, you can include images in the prompt and the model has the capability to understand that image and output a relevant response. That

is, you can include images in the prompt and the model can understand them and output a relevant response.

Aside from this, there's also other generative models with specific purpose. For image generation, popular services include DALL-E (Ramesh, et al., 2021), Midjourney (Hanna, 2023), and REimagine.ai (Gaster, 2024). AudioLDM (Oh, Kang, Moon, Choi, & Chon, 2023), Google Cloud TTS (Shekhar, 2023), and Google MusicLM (Agnostelli, et al., 2023) are leading projects for audio generation. Products such as Synthesia (Thompson, 2024), runway.AI (Germanidis, 2024), and NUWA-XL (Yin, et al., 2023) are pioneers for video generation.

Successful Examples of Ethical GenAI Implementation

Even before OpenAI's launch of GPT-3, the Bank of America has already developed their very own GenAI chatbot called Erica in 2018. Erica is a virtual powered AI assistant which has the capability to answer any bank-related inquiries and provide insights to the clients based on their spending behavior. As of April 2024, this chatbot has engaged with over 40 million customers with over 2 billion interactions. Moreover, on average, more than 98% of the clients get answers to their inquiry within 44 seconds, and if further assistance is needed, there is also an option to switch to a human agent (Aldridge, 2024). In the six years of this chatbot's existence, no major problem has been reported except for Erica's logo trademark issue (Brittain, 2024), but that is completely unrelated to its GenAI capability.

With the success of Erica, Wells Fargo, another top banking company in the US, has also launched their own virtual assistant, conveniently named Fargo, which is powered by Google's AI (Papaj, 2022). After two years, this chatbot has reached around 100 million interactions annually (Marshall, 2024). Aside from this virtual assistant, Wells Fargo has also launched another LLM called Lifesync, which provides advice to the customers for their financial goal setting and planning (Blakey, 2023), receiving a million active users in its first month (Marshall, 2024). Along with other GenAI and ML algorithms for internal use, Wells Fargo created a centralized platform for their models called Tachyon, showing their long-term vision of AI while also staying ahead among their competitors in terms of production (Marshall, 2024).

In the healthcare sector, Mayo Clinic has successfully implemented GenAI to improve patient communication and operational efficiency. They have developed an AI system called Augmented Response Technology (ART) that assists in drafting message responses to patient inquiries, which helps to save significant amounts of time for medical staff. ART has saved Mayo Clinic approximately 1,500 hours (i.e., about 2 months) per year, enhancing their ability to deliver timely and effective patient care (Cacciaglia, 2024).

In the field of journalism, the New York Times (NYT) has strategically embraced GenAI within its newsroom, guided by principles that emphasize its role as a supportive tool rather

than a replacement for human journalism. They leveraged GenAI to enhance journalistic capabilities, enabling deeper exploration of stories and broader accessibility through features like digitally voiced articles and multilingual translations (Seward, 2024). This integration underscores NYT's commitment to maintaining journalistic integrity and transparency. Every use of GenAI begins with vetted information, overseen by journalists, and undergoes rigorous editorial review to uphold the high standards of accuracy and ethics expected by NYT readers. By adhering to these principles, NYT ensures that while technology enriches their reporting, human expertise remains central to delivering trustworthy journalism in the digital age.

These examples illustrate the ethical implementation of GenAI through several key principles. Bank of America's Erica and Wells Fargo's Fargo prioritize customer satisfaction and efficiency, ensuring rapid and accurate responses while providing a seamless transition to human agents when needed (Aldridge, 2024; Marshall, 2024), reflecting a commitment to user autonomy and transparency. Additionally, these AI systems are designed to safeguard customer data, with no major incidents reported, emphasizing data privacy and security. Mayo Clinic's ART underscores ethical AI use in healthcare by enhancing operational efficiency without compromising patient care (Cacciaglia, 2024), thereby contributing to the well-being of both patients and medical staff. The New York Times exemplifies ethical journalism by using GenAI as a supportive tool rather than a replacement for human reporters (Seward, 2024), maintaining rigorous editorial standards and transparency in its AI-driven processes. Each of these implementations demonstrates how GenAI can be ethically harnessed to enhance services while upholding principles of user-centricity, data protection, transparency, and human oversight.

Lessons Learned from Ethical Failures

Along with the successful use cases are the ethical failures that resulted from the naive use of GenAI. An example of this is the recent issue about the chatbot deployed by Air Canada for customer service. The chatbot mistakenly promised a discount for a $600 ticket to a customer, but when they arrived at the airport, the airline refused to comply with that discount. Air Canada argued that the chatbot is a separate legal entity responsible for its own actions, but the tribunal rejected that argument and held them liable for that mistake (Yagoda, 2024). The airline‘s attempt to evade responsibility for the chatbot‘s mistake undermines trust and accountability in GenAI systems. Many similar cases have been reported, albeit most of them are minor in terms of media coverage and damages. Lessons learned in these kinds of ethical failures include:

1. GenAI systems are not Infallible: With the large amount of data that GenAI are usually trained from, they are prone to hallucination. GenAI hallucination is a phenomenon where the model presents inaccurate and nonsensical

information as its output. This is the reason why Air Canada chatbot has mistakenly promised a discount to the customer. Hence, it is important to acknowledge that GenAI is not always right all the time.

2. Human Oversight is Important: Given that GenAI is not 100% accurate, there is a need for a human agent to monitor and support it, especially in cases when it mistakenly outputs inaccurate information. The human agent responsible should be aware of the correct information the GenAI should have provided, or at least where to get that information from.

3. GenAI should be Transparent and Explainable: It is also important that the users are aware of how GenAI tool works and how it was able to arrive at the answer it provided. The answers provided should also include references on where it got the information from, so that users could cross-check the information provided themselves.

Public Perception and Trust

Generative AI models, once put into the mainstream, have seen a rapid adoption by the public. Within only five days of its launch in November 2022, OpenAI's ChatGPT gained one million users, becoming the fastest-growing application in history. For context, it would take other tech services like Spotify, Facebook, and Twitter more than 1,500 days (about 4 years) to reach the same milestone. However, despite the swift proliferation of ChatGPT and other GenAI services, public perception of Generative AI has not always been positive or has only been seen in Western-leaning or technologically literate countries.

A recent study investigated Twitter users' perceptions of Generative AI using over 3 million Twitter posts from 2019-2023. While most sentiments are positively correlated with exposure to AI, illustrators and artists show exceptionally negative sentiment due to concerns about unethical usage of artworks (Miyazaki, Murayama, Uchiba, An, & Kwak, 2024). A research survey answered by Computer Science students from South Korea showcased how younger users believe that GenAI can replace traditional internet searches and information foraging, while non-users tend to worry more about AI replacing people with jobs. They also understand that there is a level of distrust despite overall helpfulness, as they recognize that AI is not infallible and should be subject to human oversight (Amoozadeh, et al., 2024).

Interestingly, even with this distrust from public, other research shows that AI-generated content is more trustworthy in certain contexts. A research study compared traditional scientific summaries of 34,584 papers found in a research journal to AI-generated

summaries of the same work and compared complexity and readability through automated text analysis and participant surveys. They found that participants perceived AI-generated statements as clearer and less complex compared to human-made statements, and that the simpler statements were made by humans when in fact were AI-generated (Markowitz, 2024). Another paper suggests that news from GenAI is believed less – two experiments on 4,034 US news headlines showed that people rated those written by AI as less accurate than those written by humans. People were more likely to incorrectly rate AI-written headlines as inaccurate than human-made headlines (Longoni, Fradkin, Cian, & Pennycook, 2022).

Presenting GenAI Frameworks Across the Globe

GenAI holds potential for innovation across various sectors, but its development necessitates various ethical, legal, and cultural considerations. To address these concerns, various researchers and organizations have proposed and established GenAI frameworks, from universal to industry-specific use. This section explores prominent frameworks from across the globe to understand their diverse approaches to governing GenAI creation.

Frameworks from Countries and Organizations

There are many existing frameworks for general AI usage across the globe, often being forwarded by government agencies of countries leading in the AI and tech industry. Below are examples of some frameworks – these are not specific to GenAI but should encompass the general guidelines GenAI systems should adhere to.

Some AI regulations and guidelines worldwide are specific to certain responsible AI principles. Europe's General Data Protection Regulation (GDPR) focuses on privacy and data protection, while USA's Algorithmic Accountability Act dives deeper into transparency and fairness. China's own AI Development Guidelines tackle multiple discussions on ethics and security, while OECD's AI Principles aims to be an international standard on AI transparency, fairness, and accountability.

Singapore's Model Artificial Intelligence Government Framework, created by the country's Infocomm Media Development Authority (IMDA) and Personal Data Protection Commission (PDPC) takes a step forward from these general regulations (Personal Data Protection Commission Singapore, Infocomm Media Development Authority Singapore, 2020). Having its 2nd edition released as recently in 2019, focuses primarily on four (4) areas:

- Internal governance structures and measures: Efforts such as having a multidisciplinary body to oversee AI governance efforts, and developing standards,

guidelines, and templates to help organizations in designing and implementing AI systems responsibly

- Human involvement in AI-augmented decision-making: This includes conducting risk impact assessments and assigning categories of human involvement based on risk level (human-in-the-loop, human-over-the-loop, and human-out-of-the-loop)
- Operations Management: Acknowledges that the AI System Lifecycle has various stages and is iterative, but risk-based assessments should be conducted before any data collection or modelling
- Stakeholder interaction and communication: Emphasizes that stakeholder trust should be developed throughout the lifecycle, measures should be put in place for helping employees adapt to AI-augmented work environments, and general disclosures should be places when AI is used in products or service offerings

The 2nd edition of the ASEAN Guide on AI Governance and Ethics was released in February 2024, serving as a practical guide for organizations in the ASEAN region to design, development, and deploy traditional AI technologies in commercial and non-military or dual-use applications (Association of Southeast Asian Nations, 2024). This stands out as it focuses on fostering the interoperability of AI frameworks across administrative jurisdictions within the region. Along with guiding principles, it also touches upon the four (4) key components found in Singapore's framework.

The United States' National Institute of Standards and Technology (NIST) have also released their 1st edition of an Artificial Intelligence Risk Management Framework (AI RMF 1.0) last January 2023. Its goal is to offer a resource to organizations designing, developing, or using AI systems to help manage the many risks of AI promote trustworthy development and use of AI systems (National Institute of Standards and Technology, 2023). The document intends to be a voluntary, rights-preserving, non-sector-specific, and use-case agnostic framework, providing flexibility to organizations of all sizes and sectors throughout society. The AIM RMF Core is composed of four (4) functions:

- Govern: This establishes a culture of risk management through the system's lifecycle, involving setting up processes to identify and manage potential risks, and seeking accountability with clear roles and responsibilities.
- Map: This concentrates on understanding the AI system through characterizing its intended use and the environment it operates in.
- Measure: This focuses on assessing the likelihood and impact of potential risks identified in Map, with different quantitative and qualitative methods to evaluate the severity of risks for easier prioritization based on potential harm

- Manage: This deals with developing plans to address identified risks, like implementing controls to mitigate risks, monitoring system performance, and continuously improving risk management processes

Lastly, the European Artificial Intelligence Act (AI Act), enacted in April 2024, is the first-ever legal framework adopted by a major regulatory body, aiming to establish a foundation for trustworthy AI development and use within the European Union (The EU Artificial Intelligence Act, 2024). Its main focuses are on safety and ethics, the harmonization of standards, and the potential for global influence. The AI Act classifies AI applications into different risk categories:

- Unacceptable Risk (Prohibited): Systems deemed too risky, like social scoring for government control or subliminal manipulation and deception, are banned.
- High-Risk: Applications with significant potential for harm, such as recruitment tools and biometric ID systems, face stricter regulations to ensure fairness, transparency, and human oversight. Generative AI tools are likely to fall under this bracket.
- Low-Risk: Applications with minimal risk, such as spam filters and AI-powered games, seek minimal regulation.

Frameworks from Research Communities

To date, frameworks specifically for Generative AI systems have not been released or publicized by any major countries or organizations, but there is already an astounding number of discussions and recommendations from various research communities.

Chief editors from the Journal of Information Technology argue that it is not inherently unethical for GenAI technologies to be used for research and science. If AI is used for valid and legitimate research for which authors must take full responsibility, then it should be allowed. However, GenAI usage should also be declared in the same ways other tools are declared. They also discuss about the differing perspectives in deontological ethics (“a fair process”), where we prioritize process over outcome in moral considerations, and teleological ethics (“a good outcome”), where outcomes and results are prioritized over the nature of actions (Schlagwein & Willcocks, 2023).

Frameworks on GenAI may focus on extra ethical considerations. According to a recent article on conceptual frameworks (Zlateva, Steshina, Petukhov, & Velev, 2024), this may be towards general GenAI systems, wherein you consider five blocks:

- Awareness: Recognizing ethical issues and knowing potential limits of GenAI

- Evaluation: User analysis for those affected by GenAI results, risk analysis for finding magnitude, and benefit assessment of GenAI capabilities compared to risks
- Decision-Making: Ethical guidelines to establish clear rules on widely accepted ethical principles
- Implementation: Ethics in design, transparency, explainability, and user feedback
- Review Process: Results to be monitored consistently and framework to be reviewed periodically

Large Language Models (LLMs) are arguably the face of GenAI models, which may explain the pervasiveness of LLM-specific regulatory frameworks. For example, a recent editorial piece published in *Nature* stresses the need for regulatory oversight of LLMs and other GenAI tools in healthcare, as they posit that they need different regulations from existing AI-based technologies – some reasons are due to scale and complexity, hardware requirements, broader applicability, real-time adaptation, and data privacy and security (Mesko & Topol, 2023).

Another article suggests that differentiated terminologies are needed for large GenAI models to allocate regulatory duties to specific actors and activities in the GenAI value chain (Hacker, Engel, & Mauer, 2023). They specifically recommend three rule layers:

1. Rule Layer 1: Existing, technology-neutral regulations such as GDPR or non-discrimination provisions
2. Rule Layer 2: Single out high-risk applications and *not* pre-trained models, as it is infeasible to compel LLM developers to draw up comprehensive risk management systems
3. Rule Layer 3: Mandate collaboration between actors in the GenAI value chain for compliance purposes

Frameworks under education have been more prevalent, perhaps due to the more obvious usage and potential risks of GenAI systems being used by teachers and learners. One study proposes a GenAI ethical foundational principles in teachers' education framework (GENAIEF-TE), which focuses on (Radwan & McGinty, 2023):

- Transparent Accountability: provide all stakeholders (e.g. educators, students, administrators) with comprehensive info on functioning of GenAI systems (e.g. algorithmic operations, data usage, decision-making processes). In practice, this

means using GenAI tools for student assessment and providing students with explanation on how these tools evaluate their work.

- Privacy and Secure Data Management: Developers engage with stakeholders to make informed decisions on GenAI deployment in classrooms, including not disclosing sensitive information or perpetuating biases, or dealing with "emotional" GenAI technologies.
- Culturally Sensitive and Inclusive Fairness: GenAI systems acknowledge, respect, and accommodate diverse cultural, social, and ethical norms of users. Fairness in GenAI means that their outputs do not favor nor disadvantage specific groups based on cultural or other demographic variables.
- Community Centered Design: GenAI tools are also contextually relevant to the educational communities it serves
- Transparent Data and Algorithmic Literacy: Stakeholders should possess a basic understanding of data-related concepts, like how data is gathered, processed, and applied within GenAI systems.
- Pedagogy-Centered Design: Intertwine educational theories and practices in the development and operation of GenAI tech.

Another framework proposes an Artificial Intelligence Assessment Scale (AIAS) for the ethical integration of Generative AI in Educational Assessment (Perkins, Furze, Roe, & MacVaugh, 2024). It proposes five (5) levels:

1. No AI: AI must not be used at any point during assessment
2. AI-Assisted Idea Generation and Structuring: No AI content is allowed in the final submission
3. AI-Assisted Editing: AI can be used, but your original work with no AI content must be provided in an appendix
4. AI Task Completion, Human Evaluation: You will use AI to complete specified tasks in your assessment. Any AI created content must be cited.
5. Full AI: You may use AI throughout your assessment to support your own work, and do not have to specify which content is AI generated

What is Still Needed?

As seen in literature review, most frameworks related to Generative AI tools and implementations are on general purpose and traditional AI systems. While these have significant merit, there is still a need for clarity on whether Generative AI systems and tools should be held under the same guidelines and regulations, or if additional and more intricate measures should be suggested.

It is our belief as Aboitiz Data Innovation (ADI) that there is still a need for separate terminologies, objectives, and action points for generative AI-based systems, due to its unique nature of generating content rather than predicting. By just providing a few words or audio as input, high-quality content can be easily generated with the help of GenAI tools. From this functionality, various practical uses involve conversational chatbot, coding assistant, content creation, machine translation, etc. However, alongside the great quality of life that these tools provide are the risks and uncertainty that they may bring to our society. According to identity verification company Onfido, fraudsters are utilizing GenAI to generate deepfakes for purposes like business fraud, identity theft, and social engineering attacks (Takruri, 2023).

We identify a need for a framework that could govern these new technologies to adhere to human ethics and laws. The GenAI Governance Framework developed by Aboitiz Data Innovation addresses this critical need by offering a comprehensive, lifecycle-based approach. We list down below why a framework is essential for the responsible development and deployment of GenAI:

1. Lifecycle-Based Focus

   Ethical principles are crucial throughout GenAI development, but their application varies at different stages. Our framework goes beyond mere principles. It delves into each stage of the GenAI lifecycle – from problem definition to ongoing monitoring – highlighting the critical principles that require specific focus at each juncture. This actionable guidance empowers developers and organizations to translate ethical principles into practical strategies, ensuring responsible development and deployment.

2. Mitigating Risks Across Industries

   GenAI applications span diverse domains, each presenting unique risks and ethical considerations. Our framework acknowledges this by incorporating beneficial and harmful examples across various industry use cases. This real-world context allows stakeholders to identify potential risks and opportunities specific to their domain, fostering a nuanced and stricter approach compared to general-purpose AI. With exponentially greater potential risks in GenAI usage, our framework equips

developers to mitigate those risks upfront, considering the specific societal and ethical implications each use case presents.

3. Framework vs. No Framework: A Clear Distinction in Outcomes

   The GenAI Governance Framework offers a stark contrast to the potential pitfalls of GenAI development without a structured approach. Here's a breakdown of the benefits and risks associated with each scenario:

   - GenAI with Framework:
     - Benefits: Reduced bias, improved transparency, mitigated safety risks, responsible data usage, ethical alignment, and demonstrably positive societal impact.
     - Risks: Increased development time due to adherence to ethical considerations, potential need for revisions based on ongoing monitoring.
   - GenAI without Framework:
     - Benefits: Potentially faster development timeline (initially).
     - Risks: Unintended bias, lack of transparency, potential safety vulnerabilities, privacy violations, unethical applications, and negative societal consequences.

By adopting the GenAI Governance Framework, stakeholders can navigate the complexities of GenAI development and mitigate these risks. Investing in the initial stages through a framework-based approach avoids costly problems and ethical downfalls later in the development and deployment process.

The Proposed Framework

The responsible development, deployment, and governance of Generative AI systems necessitates a comprehensive governance framework that addresses ethical considerations throughout the entire lifecycle, and that goes beyond simply listing ethical principles. This paper proposes a Generative AI Governance Framework modeled after the six stages of GenAI system development and eleven key principles.

This Framework is built based on the implementation and deployment of AI and GenAI-related systems, found through both extensive research and our experience as a leading software company in Southeast Asia. ADI specializes in cutting-edge data science

and AI solutions, with key capabilities in data management & architecture, smart city advisory, and Generative AI. Our GenAI for Enterprises service has allowed different organizations to automate their data management through intelligent search and chatbots, while our Generative AI Lab enables enterprises to create innovation hubs for GenAI-backed exploration and development. We have worked with organizations in and outside the Aboitiz conglomerate consisting of energy, construction, financial services, real estate, and smart cities industries – we utilize this experience to ensure the principles we outline are actionable and directly address the challenges encountered throughout the GenAI lifecycle.Governance Principles

A generative AI-based governance framework should aim to create a balanced and comprehensive approach to managing the risks and benefits associated with generative AI, ensuring its development and use are aligned with societal values and ethical standards. It is multi-faceted that deals primarily to ensure the responsible and beneficial use of the technologies. The key considerations common across the various governance frameworks mainly emphasizes the following objectives:

1. Ethical Use, to ensure that generative AI is developed and used in ways that align with ethical principles;
2. Bias and discrimination mitigation, to identify, address, and mitigate biases in AI models and their outputs, preventing discriminatory practices and promoting equality;
3. Transparency and explainability, to promote transparency in how generative AI models work and make decisions;
4. Data protection and privacy, to safeguard personal data and privacy rights, ensuring that generative AI systems comply with data protection laws and ethical standards;
5. Accountability and responsibility, to establish clear lines of accountability and responsibility for the actions and outputs of generative AI systems, ensuring that there are mechanisms for addressing harm or errors;
6. Safety and reliability, to ensure that generative AI systems are safe, reliable, and secure, preventing misuse and unintended harmful consequences;
7. Regulatory compliance, to ensure that the development and use of generative AI comply with relevant laws, regulations, and industry standards;
8. Social impact, to consider the broader social implications of generative AI, ensuring that its development and deployment contribute positively to society and do not harm public interests;

9. Innovation and advancement, to foster innovation while balancing it with ethical considerations, ensuring that generative AI technologies advance in ways that are socially and ethically responsible;

10. Stakeholder engagement, to involve a diverse range of stakeholders, including policymakers, industry experts, and the public, in the governance process to ensure that multi-perspectives are considered; and

11. Education and training, to train and inform developers, users, and stakeholders on the ethical use and limitations of generative AI, promoting informed and responsible use.

These principles were based on existing frameworks found in the previous section. In Figure 1, we observe the eleven key principles grouped into three (3) broad categories for easier reference. These categories are:

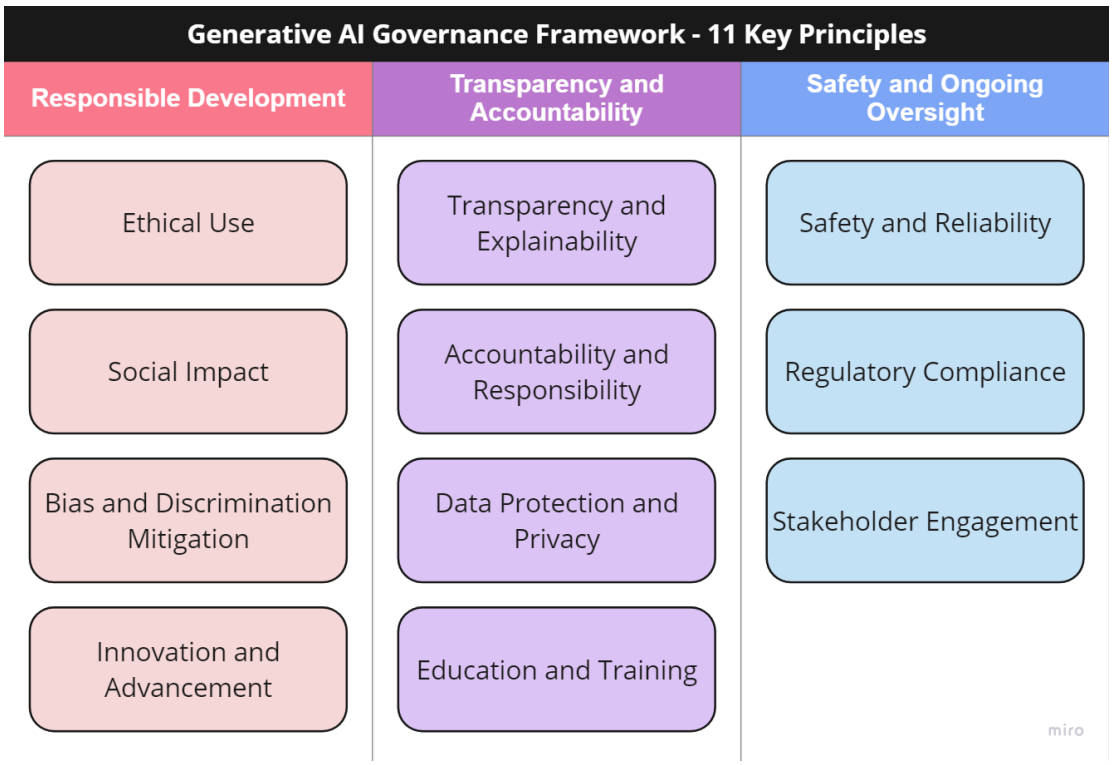


Figure 1. 11 Key Principles of the Generative AI Governance Framework

The GenAI system development lifecycle

The GenAI system development lifecycle places a stronger emphasis on data quality, utilizing specialized model architecture, and prioritizing creative output and safety considerations during evaluation and deployment.

1. Problem Definition: This is like the process for problem definition concerning the potential use of a traditional AI system. Stakeholders should identify the creative task or content generation goal for the Generative AI system. Consider if the problem could also be solved through non-AI means (i.e. process improvement or other technical services), or through non-generative means.

2. Data Acquisitioning and Preprocessing In a traditional AI system, relevant and non-biased data is also crucial for training the AI model, though it may be easier to process and extract due to commonly available techniques. The quality and diversity of data is more crucial for Generative AI, as large amounts of high-quality data relevant to the desired output (text, images, code, video, audio, etc.) are needed to train the model effectively. Techniques such as data augmentation may be used to create more data training variations.

3. Model Development and Training: There are plenty of suitable traditional AI model architectures that have been researched and deployed; a trait that GenAI system lacks due to its relative infancy. GenAI systems often employ specialized model architectures like GANs or transformers designed for content generation. Training may involve iterative refinement to ensure the generated content aligns with desired style and avoiding nonsensical outputs.

4. Model Evaluation and Refinement: Model performance of traditional AI systems is assessed and refined, but metrics and evaluation techniques are already present and are easily explainable and attributable to real-life implications such as revenue or productivity. Since GenAI outputs are more complicated than typical AI outputs (e.g. binary outputs or clustering), different evaluation techniques and manual annotation may be needed in this phase. Safety checks are also crucial to identify potential biases or offensive content that is generated.

5. Deployment: Deploying traditional AI models to a production environment for real-world use is a prevalent operational step that most organizations are already engaging in. There are also companies and products which specialize in MLOps or AI-Ops services. Deployment of GenAI systems may involve a more controlled rollout instead of releasing a full product immediately, especially for high-risk applications. Constant feedback and control loops along with human oversight are necessary to ensure the quality and safety of generated content before and as it is released to the public.

6. Monitoring and Maintenance: Regular AI models are also continuously monitored for performance. However, monitoring for potential biases, safety issues, or unintended consequences as its being used and operationalized is more essential for GenAI systems due to the lack of standardized protocols or mitigation strategies. Feedback loops and transparent updates should be established to improve the model's effectiveness and ensure responsible use.

The lifecycle of a Generative AI system may share many similarities with a traditional or general-purpose AI system, but the key differences lie in data quality and additional human oversight, due to the scalability and added potential risks an GenAI system poses. We visualize this lifecycle in Figure 2.

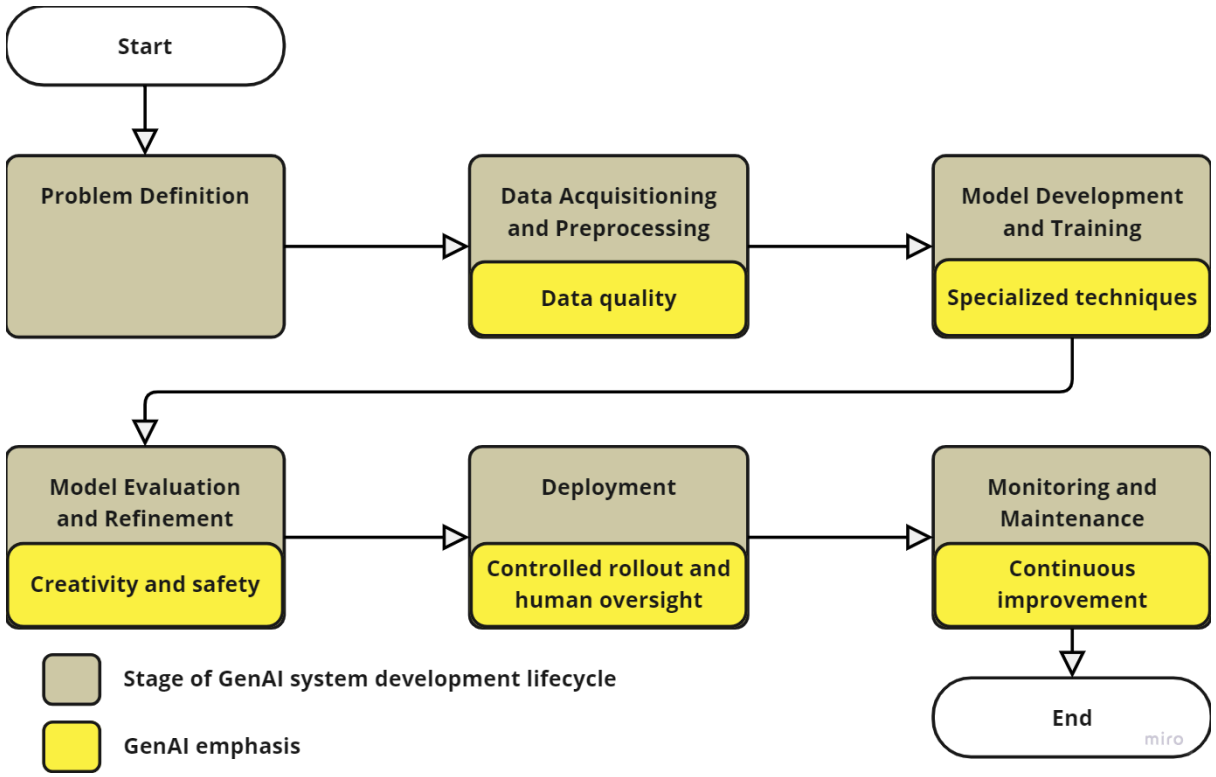


Figure 2. GenAI System Development Lifecycle

Governance throughout each stage

As each GenAI governance principle and lifecycle stage has been defined, we now delve into each stage exploring how each principle applies. In Table 1, we provide considerations and real-world examples of both responsible and irresponsible practices. By examining these considerations, stakeholders can now translate the framework's principles into responsible action. These principles were defined based on Aboitiz Data Innovation's experience with various general-purpose AI and GenAI projects, and through the use cases we've researched and analyzed.

Stage	Principles	Considerations
Problem Definition	• Ethical Use • Social Impact • Stakeholder Engagement	• Clearly define purpose and intended use of GenAI system • Consider potential social implications and benefits with risks

		• Identify relevant stakeholders and involve them in early discussions
Data Acquisition and Preprocessing	• Bias and Discrimination Mitigation • Data Protection and Privacy • Transparency and Explainability	• Source data responsibly and ethically, considering potential biases • Implement robust data anonymization and privacy protection measures • Document data sources and preprocessing methods for transparency
Model Development and Training	• Bias and Discrimination Mitigation • Transparency and Explainability • Accountability and Responsibility	• Choose training data and algorithms that minimize bias • Develop methods to explain model outputs and decision-making processes • Clearly define roles and responsibilities for model development and deployment
Model Evaluation and Refinement	• Bias and Discrimination Mitigation • Transparency and Explainability • Safety and Reliability	• Evaluate models for potential biases and fairness in outputs • Develop methods for assessing model explainability and interpretability • Test for potential safety risks and unintended consequences of model outputs
Deployment	• Ethical Use • Transparency and Explainability • Accountability and Responsibility • Regulatory Compliance	• Deploy the model in a way that aligns with its intended purpose and ethical principles • Ensure transparency regarding use of GenAI in deployed system • Clearly define lines of accountability for model performance and potential harm • Comply with all relevant regulations regarding GenAI deployment and data usage
Monitoring and Maintenance	• Bias and Discrimination Migration • Safety and Reliability • Transparency and Explainability • Accountability and Responsibility	• Continuously monitor model performance to detect and address bias drift • Monitor for safety issues and potential vulnerabilities in the deployed system • Maintain transparency regarding model updates • Be prepared to address any harm caused by the model and take corrective actions

		• Establish committees or approval boards where monitoring information is either reported or available to

Table 1. The six-stage GenAI system development lifecycle and corresponding GenAI Governance principles

This framework integrates the identified principles into the six-stage Generative AI system development lifecycle. Other principles, such as Innovation and Advancement, Stakeholder Engagement, and Education and Training, may be relevant throughout the entire lifecycle, as we foster a culture of continuous learning and responsible innovation. We further detail each principle on each stage below, showcasing good and bad implementations of each principle.

Problem Definition

1. Ethical Use: This emphasizes that the intended purposes and application of the GenAI system should align with the organization or jurisdiction's ethical values. Developers and business stakeholders should consider potential risks and benefits to society, fairness, and human wellbeing

    a. Beneficial Example: An elementary school develops a GenAI system to create personalized learning materials that cater to individual student needs and learning styles, promoting educational equity.

    b. Harmful Example: A media agency develops a GenAI system to generate deepfakes for entertainment purposes without considering potential misuse for political manipulation or defamation.

2. Social Impact: Consider the broader societal implications of the GenAI system, meaning if unintended users were able to use the features of your system for other purposes. Analyze potential positive and negative impacts on areas like employment, privacy, and social discourse.

    a. Beneficial Example: A research institute develops a GenAI chatbot to pair with an existing model that analyzes climate data and predicts extreme weather events. The goal is to improve preparedness and mitigate potential social and economic damage from natural disasters.

    b. Harmful Example: A social media platform creates a GenAI system to personalize newsfeeds, but the algorithm inadvertently creates echo chambers that amplify misinformation and polarize public discourse.

3. Stakeholder Engagement: This highlights the importance of involving relevant stakeholders in discussions about the GenAI system's development and use, even before any model is developed. This could include researchers, policymakers, ethicists, and potentially the public.

    a. Beneficial Example: A city government collaborates with AI developers and citizen groups to develop a GenAI system for optimizing traffic flow, by generating simulated data and using a chatbot for explainability. Stakeholder engagement ensures the system addresses public concerns and promotes equitable traffic management.

    b. Harmful Example: A company develops a GenAI system for summarizing applicant application data and automating hiring decisions without involving HR professionals or seeking input from potential job candidates. This can lead to biased hiring practices that exclude qualified individuals.

Data Acquisition and Preprocessing

1. Bias and Discrimination Mitigation: This showcases the need to identify and address potential biases within the data used to train the GenAI system. Biased data can lead to discriminatory outputs that perpetuate societal inequalities.

    a. Beneficial Example: A healthcare company sources data from diverse populations for a GenAI system that assists their doctors in diagnoses. The data is preprocessed to mitigate biases based on race, gender, or socioeconomic background.

    b. Harmful Example: A financial institution trains a GenAI system for loan approvals using synthetic data that reflects existing biases in lending practices, leading to discriminatory loan denials for certain demographics.

2. Data Protection and Privacy: This principle ensures that user information collected for training the GenAI system is handled responsibly and adheres to data protection regulations. Developers and deployment teams should implement anonymization and security measures to safeguard privacy.

    a. Beneficial Example: A research team anonymizes patient data before using it to train a GenAI system for drug discovery and explanation. This protects patient privacy while allowing for valuable medical research.

    b. Harmful Example: A social media company collects user data without explicit consent, using it to train a GenAI system that generates personalized

advertisements. This raises privacy concerns and potential violations of data protection laws.

3. Transparency and Explainability: This emphasizes documenting data sources and preprocessing methods used to prepare data for GenAI training in ways that both AI experts and non-AI experts cna understand. This transparency allows for scruiny and helps identify potential biases or data quality issues.

    a. Beneficial Example: Researchers publish a paper detailing data sources and preprocessing techniques used to train a GenAI system for facial recognition, allowing for peer review and replication of research findings.

    b. Harmful Example: A company develops a GenAI system which helps users in financial trading, without disclosing the source or processing methods of the training data. This lack of transparency raises concerns about potential biases and model manipulation, i.e. preferring certain stocks or commodities.

Model Development and Training

1. Bias and Discrimination Mitigation: This focuses on selecting training data and algorithms that minimize potential biases in the GenAI model. Techniques such as debiasing algorithms or data augmentation can help address this concern.

    a. Beneficial Example: A face creation system for a video game is trained on a dataset balanced across various ethnicities, ages, and genders. This helps mitigate potential biases in facial recognition accuracy across different demographics.

    b. Harmful Example: A recruitment agency uses a GenAI system for resume screening trained on historical data reflecting existing hiring biases. The GenAI system explains the rationale on answers but can perpetuate discrimination against certain demographics in the hiring process.

2. Transparency and Explainability: This highlights the importance of developing methods to explain the GenAI model's outputs and decision-making processes. This can involve techniques like interpretable AI models or feature/text attribution methods.

    a. Beneficial Example: A travel itinerary chatbot powered by GenAI provides explanations for location decisions, allowing human reviewers to understand the model's reasoning and identify potential biases.

    b. Harmful Example: A social media platform deploys a GenAI system that curates user feeds without explaining how content is selected nor ranked.

This lack of transparency hinders user understanding and trust in the algorithm's platforms.

3. Accountability and Responsibility: This establishes clear lines of accountability and responsibility for the development of the GenAI model before it gets to testing and deployment. This includes data scientists, engineers, and project managers.

    a. Beneficial Example: A research team developing a GenAI system for image generation clearly defines roles and responsibilities for model development, sourcing data, safety testing, and potential liability in case of inappropriate generation.

    b. Harmful Example: A company deploys a GenAI system for content moderation in a Discord server without clear lines of accountability for biased or unfair content moderation decisions.

Model Evaluation and Refinement

1. Bias and Discrimination Mitigation: This emphasizes evaluating the GenAI mode for potential biases and fairness in its outputs. Beyond accuracy, fairness metrics that measure differential performance across groups or categories are crucial.

    a. Beneficial Example: A healthcare GenAI system for cancer detection and explainability is rigorously evaluated for potential biases in diagnoses across different patient demographics.

    b. Harmful Example: A criminal justice system deploys a GenAI system to summarize lengthy documents and create risk assessment profiles, without evaluating potential biases in predicting recidivism rates. This can lead to discriminatory sentencing outcomes.

2. Transparency and Explainability: This focuses on assessing the model's explainability and interpretability. This allows for identifying potential biases and ensures decisions are not based on "black-box" outputs.

    a. Beneficial Example: Researchers develop a GenAI system for summarizing scientific documents and provides visualization tools to understand how the model assigns keywords and identifies relationships between concepts.

    b. Harmful Example: A political campaign deploys a GenAI system for micro-targeting and personalizing messages to voters, without disclosing how the model identifies and targets specific demographics. This lack fo transparency raises concerns about manipulation and voter privacy.

3. Safety and Reliability: This stresses the need of testing for potential safety risks and unintended consequences of the GenAI model's outputs. This includes potential misuse or vulnerabilities in generated contents.

    a. Beneficial Example: A tech company developing a GenAI system for generating news articles performs rigorous tests to ensure factual accuracy and prevent the creation of misleading or harmful content.

    b. Harmful Example: A financial institution mistakenly links a database with PIIs and user passwords to their GenAI chatbot's knowledge base, where users could ask the chatbot for passwords without alerts.

Deployment

1. Ethical Use: This ensures that the GenAI system is deployed in a way that aligns with its intended purpose and ethical principles defined during problem definition.

    a. Beneficial Example: A retail company deploys a GenAI system for generating product descriptions with clear disclaimers about AI-generated content and human oversight in the review process.

    b. Harmful Example: A government agency deploys a GenAI system for social surveillance, creating features attributing to good or bad social standing, without clear guidelines and public oversight.

2. Transparency and Explainability: This emphasizes informing users about the use of GenAI in the deployed system and the limitations of generated content. Transparency builds trust and allows users to make informed decisions.

    a. Beneficial Example: A music streaming service utilizes a GenAI system to create mashups of certain songs in a playlist but discloses to users and song creators that these are generated by AI algorithms.

    b. Harmful Example: A political party uses a GenAI system to generate false social media content about political opponents, without disclosing the use of AI or the potential for manipulated generated content.

3. Accountability and Responsibility: This focuses on clearly defining lines of accountability for potential harm and inappropriate outputs caused by the deployed GenAI system. This could involve establishing human oversight mechanisms and clear channels for reporting issues.

    a. Beneficial Example: A financial institution deploying a GenAI system for customer service complaints establishes a clear escalation process for users

who want to talk to humans or are uncomfortable with the system's responses.

b. Harmful Example: An airline company does not take responsibility for any false discounts or promotions their chatbot communicates to customers.

4. Regulatory Compliance: This looks at ensuring compliance with all relevant regulations regarding GenAI deployment and data usage. This includes data privacy regulations, algorithmic fairness laws, or specific industry regulations.

    a. Beneficial Example: A European healthcare agency deploying a GenAI system for summarizing patient documents ensures compliance with data privacy regulations like GDPR and the new AI Act, adhering to ethical guidelines for AI in healthcare.

    b. Harmful Example: A photo editing software now retroactively claims ownership of all files uploaded on the software to now be used for training and deployment of their live GenAI editing tool.

Monitoring and Maintenance

1. Bias and Discrimination Mitigation: This emphasizes continuously monitoring the GenAI system for potential bias drift, which can occur over time as data or usage patterns change.

    a. Beneficial Example: A company using GenAI software for summarizing legal documents has alerts for bias drifts and regularly updates the training data to mitigate identified biases.

    b. Harmful Example: A tech company that made a GenAI video generation service fails to detect inappropriate and offensive content being made with its service.

2. Safety and Reliability: This evaluates for vulnerabilities in generated content or potential misuse of the system, continuously monitoring the deployed GenAI system for safety risks and unintended consequences.

    a. Beneficial Example: An economic think tank monitors a GenAI system used for generating financial reports for anomalies or potential manipulation of financial data by the model.

    b. Harmful Example: A social media platform fails to monitor its GenAI system for generating summarized content in group chats, allowing the spread of

misinformation or harmful content that can negatively impact group discourse.

3. Transparency and Explainability: This maintains transparency regarding model updates and changes made to the GenAI system during ongoing monitoring and maintenance, that are understandable to both AI and non-AI experts.

    a. Beneficial Example: A video game company developing a game with procedurally generated world environments informs users about updates made to the GenAI model for improved performance and user experience.

    b. Harmful Example: A tech company that launched image generation software does not disclose that new training data was taken from a database of artists' work without their consent.

4. Accountability and Responsibility: This emphasizes ongoing responsibility for deployed GenAI system's performance and potential harm, including adapting accountability structures as the system evolves and becomes more complex.

    a. Beneficial Example: A research team developing a GenAI system for voice generation continuously monitors the model's accuracy and takes responsibility for avoiding deepfakes or inaccurate biometric identification.

    b. Harmful Example: A research team that develops a GenAI system for voice generation abandons their responsibility when a financial institution using their service has data breaches due to their system.

Ethics of GenAI with Several Industries through the lens of the GenAI Governance Framework

This section delves into ethical considerations specific to various sectors actively integrating GenAI. We explore the intersection of GenAI with existing ethical frameworks within each domain and through the perspective of our proposed GenAI Governance Framework.

Education Sector

With the emergence of GenAI, the academe has been one of the most affected sectors, especially for cases where students are maliciously using GenAI to generate content for paper submissions in their respective classes. According to the Turnitin company, 22 million papers are suspected to be AI generated in the past year for all the papers that passed through their system, with a performance of less than 1 percent false positive rate for full

(Hoover, 2024). However, given a significant error rate, which would likely blow up in cases of short papers (e.g., one-page essays), there is still a dilemma whether to fully trust the results of such AI detection tools for flagging AI generated papers. Moreover, a study (Liang, Yuksekgonul, Mao, Wu, & Zou, 2023) argued that AI detectors are biased against non-native English writers, with a 61.3% false positive rate, thus should not be the only basis for flagging AI generated papers.

However, should all AI generated content be banned in all academe-related works? According to some, the use of GenAI in research and science is not inherently unethical, given that the tools are used for valid and legitimate research for which authors must take full responsibility. For such cases, the GenAI tools must also be properly declared like the other tools used in the research. This can be a basis for a framework proposal that can be suggested to be implemented in the way of working of the academe around the world (Schlagwein & Willcocks, 2023).

A study proposes a GenAI ethical foundation principles in teachers' education (GENAIEF-TE) framework, which focuses on a transparent accountability to educators, administrators, and students; a privacy and secure data management; a culturally sensitive and inclusive fairness to all; a community-centered design; a transparent data and algorithmic literacy; and a pedagogy-centered design. With this framework, educators can maximize the advantage of GenAI and effectively integrate it in their teaching methods. In this way, the educators will be able to guide the way that their students use GenAI, rather than it being used secretively (Radwan & McGinty, 2023).

In support of this approach, some universities have taken strides, such as Maastricht University in the Netherlands. Maastricht University exemplifies a comprehensive integration of GenAI in academic settings, aligning closely with the principles advocated in frameworks such as the GENAIEF-TE. Maastricht University leverages ChatGPT to enhance educational practices, particularly in problem-based learning (PBL) environments. By integrating ChatGPT, which can generate accurate and readable responses based on vast knowledge databases, the university fosters opportunities for students and educators to explore, critique, and refine AI-generated outputs within their learning processes (Maastrict University, 2024). This approach not only supports digital and AI literacy but also encourages critical thinking and reflective learning among students, aligning with Maastricht University's commitment to constructive, contextual, collaborative, and self-directed learning principles in PBL.

Moreover, Maastricht University has developed its own guidelines for the ethical and effective use of ChatGPT in PBL environments (Maastricht University, 2023). These guidelines ensure that students engage with AI-generated content in a transparent and responsible manner, promoting integrity in academic practices while harnessing the educational potential of AI technologies.

Additionally, the proposed framework can guide the ethical use of GenAI in education through specific implementations:

a. Transparent Accountability: Educators and institutions must provide comprehensive information to stakeholders about the functioning of GenAI systems. This includes explaining how these tools evaluate student work, which can help in demystifying the technology and promoting trust.

b. Privacy and Secure Data Management: Schools should ensure that any deployment of GenAI respects the privacy of students and staff, avoiding the disclosure of sensitive information and biases.

c. Culturally Sensitive and Inclusive Fairness: AI systems should be designed to respect and accommodate diverse cultural and social norms, ensuring fairness and avoiding disadvantaging any group.

d. Community-Centered Design: GenAI tools should be beneficial to the educational communities they serve, enhancing rather than disrupting existing educational practices.

e. Pedagogy-Centered Design: AI technologies should align with educational theories and practices, supporting teaching and learning goals effectively.

Manufacturing Sector

GenAI offers various quality of life improvements in manufacturing by streamlining processes, optimizing resource utilization, and implementing advanced quality control measures (Doanh, et al., 2023). For instance, when it comes to product design, given enough design blueprints as the dataset, we can create a GenAI system that can automate the designing process by acquiring the knowledge of patterns and correlations among the past designs (Doanh, et al., 2023). With this kind of automation, people involved in the design process can explore and ideate designs more effectively. The same idea holds true for other use cases of GenAI within the manufacturing pipeline (Takyar, 2024).

However, with the quality of work that these automations provide, manufacturing companies may deem some workers to be unnecessary, leading to a mass layoff and loss of livelihood. Is it ethical for companies to replace human workers with automation? Unfortunately, the logical answer to this is yes since automation is a tool for humanitarian aid, but human agents are still essential in this process (Dourado, 2021). GenAI is not completely infallible, and it can operate without human supervision. A simple hallucination of GenAI may result in huge profit loss, which is critical in manufacturing. Therefore, what the workers can do to avoid losing their job in this technological age is to upskill in the use of GenAI in their workflow.

The proposed framework provides a structured approach to ensure the ethical use of GenAI in manufacturing, emphasizing accountability, safety, and stakeholder engagement.

a. Transparent Accountability: Manufacturers should maintain transparency about how GenAI systems are used in production processes, including any decisions made by these systems.

b. Safety and Reliability: Continuous monitoring for vulnerabilities in GenAI-generated content can prevent misuse and ensure safety in manufacturing environments.

c. Community and Worker Engagement: Engaging with employees about the deployment and impact of GenAI tools can help in adapting these technologies effectively while maintaining trust and morale among workers.

By adhering to these guidelines, manufacturers can harness the potential of GenAI for innovations in product design, predictive maintenance, and supply chain optimization, while also ensuring ethical and safe practices.

Energy Sector

Numerous companies in the energy sector are prioritizing the development and deployment of GenAI as a strategic imperative (Velasco, 2024). This trend reflects a growing recognition of GenAI's transformative potential in optimizing operations, enhancing efficiency, and supporting sustainability goals across the industry. Example real-life use cases of GenAI in this industry are:

a. Load Forecasting: A major electricity distributor in eastern Turkey implemented generative AI, specifically a Generative Adversarial Network (GAN), to enhance load forecasting accuracy. The AI model successfully captured complex, non-linear patterns in electricity consumption data, outperforming traditional time series methods and significantly improving forecast precision (Avci, 2023).

b. Power Outage Prediction: In southern Turkey, a utility company utilized generative AI to predict power outages by training a GAN model on historical weather data and past outage incidents. This approach proved superior to traditional ML methods previously employed, enabling the company to accurately forecast outage locations and severity. As a result, resource allocation for outage management became more efficient, minimizing downtime and enhancing customer satisfaction (Avci, 2023).

c. Preventive Maintenance: In northern Turkey, an electricity distribution company applied generative AI for proactive maintenance of distribution equipment. Using a GAN model trained on equipment data, the company simulated potential

degradation scenarios, enabling timely preventive interventions. This approach reduced unexpected failures, enhancing overall system reliability (Avci, 2023).

Incorporating GenAI into the energy sector not only enhances operational efficiency but also introduces significant ethical considerations that must be addressed. Hence, in addition to the innovation involved, a policy that will serve as an ethical framework is also essential for this seismic shift toward GenAI (Velasco, 2024). With the proposed framework, we can ensure that the integration of GenAI aligns with ethical standards and societal benefits through the following implementations:

a. Sustainability and Environmental Impact: GenAI tools should be used to promote sustainability, such as optimizing energy consumption and improving the efficiency of renewable energy sources.

b. Accountability and Responsibility: Energy companies should take responsibility for the environmental and societal impacts of GenAI applications, ensuring they contribute positively to global sustainability goals.

c. Stakeholder Engagement: Continuous dialogue with stakeholders, including consumers and regulatory bodies, is crucial for accountable use of GenAI in energy systems.

Banking Sector

GenAI is increasingly revolutionizing the banking sector, offering significant opportunities for efficiency gains and innovation across various operational facets. Top consulting firm McKinsey & Company highlights how GenAI enhances customer interaction through advanced chatbot functionalities, improves fraud detection capabilities, and automates complex tasks such as code development and regulatory reporting (Buehler, et al., 2024). It has also been noted that major financial institutions utilize GenAI to enhance credit risk evaluations by analyzing extensive datasets encompassing credit history, income levels, and employment status (Barde & Kulkarni, 2023). These AI-driven systems have demonstrated significant reductions in false positives by 40-75% and improved fraud detection rates by 50-80%, underscoring GenAI's pivotal role in bolstering operational efficiency and risk management strategies within banking (Barde & Kulkarni, 2023).

Despite these advantages, the integration of GenAI in banking raises critical ethical considerations that must be carefully addressed to mitigate potential risks. With the

proposed framework as guide, we can address the potential risks through the following implementations:

a. Transparency and Explainability: Banks must ensure that AI systems are transparent and their decisions explainable to customers and regulators. This helps in maintaining trust and compliance with regulatory requirements.

b. Privacy and Security: Protecting customer data is paramount. GenAI systems should be designed to uphold the highest standards of data privacy and security, preventing unauthorized access and misuse.

c. Fairness and Non-Discrimination: AI systems should be rigorously tested to ensure they do not perpetuate biases or discriminate against any group of customers.

By adhering to these principles, banks can enhance their services and operational efficiency while maintaining ethical standards and customer trust. These integrations across different sectors demonstrate how the proposed framework can guide the responsible and beneficial use of GenAI, ensuring that its deployment aligns with ethical principles and societal values.

Considerations when Implementing the GenAI Governance Framework

Translating the principles and recommended steps found in the Generative AI Governance Framework into practical action requires careful consideration. This section dives into key considerations that stakeholders must address when implementing the framework into their specific contexts. This serves as a practical guide for researchers developing new GenAI models, organizations integrating GenAI into their workflow, or policymakers crafting regulations for GenAI.

Stakeholder Engagement and Communication

The successful implementation of the Framework requires a collaborative approach – this section delves into strategies for fostering robust communication and collaboration among stakeholders across all stages of the GenAI development lifecycle.

Engaging with Stakeholders

Identifying key stakeholders

In the typical GenAI system development lifecycle, there is a wide range of key stakeholders involved. Depending on the system and stage of development, some of these stakeholders may be more directly entangled than others.

1. Developers and Researchers
    a. Data Scientists and Machine Learning Engineers: They are responsible for building and training the GenAI model, ensuring technical proficiency and adherence to ethical principles.
    b. AI Researchers: They conduct research on the capabilities and limitations of GenAI, contributing to advancements in the field while considering ethical implications.
2. Businesses and Organizations
    a. Product Managers, Infrastructure Teams, and Company Executives: They drive the development and deployment of GenAI systems within their organizations, ensuring alignment with business goals.
    b. Legal and Compliance Teams: They advise on regulatory compliance for GenAI deployment and data usage, safeguarding legal adherence.
3. Regulatory Bodies and Policymakers
    a. Government Agencies: They develop and implement policies and regulations for GenAI development and deployment, balancing innovation with public safety and other ethical considerations.
    b. Standardization Organizations: They establish standards for responsible AI development and data usage, promoting best practices and industry-wide ethical implementation.
4. End Users and Consumers
    a. Individuals interacting with GenAI system: They have a right to understand how GenAI is used with its potential limitations, empowering them to make informed decisions whether to use the system, and raise concerns about potential manipulations or biases.
    b. General Public: They hold a stake in the responsible development and deployment of GenAI – public awareness and education is crucial for informed discussions and trust-building

Depending on the specific GenAI system being developed, additional stakeholders might be involved, such as domain experts, investors, ethicists, public interest groups, or specific communities impacted by the technology.

Strategies for effective engagement

The successful and responsible development of a GenAI system necessitates a collaborative approach throughout the entire lifecycle. In Table 2, we detail the different responsibilities and strategies each stakeholder group can employ throughout the six stages of GenAI development.

Stage \ Stakeholder Group	Developers and Researchers	Businesses and Organizations	Regulatory Bodies and Policymakers	End Users and Consumers
Problem Definition	• Leverage technical expertise to identify potential applications of GenAI and define the specific problem the system aims to address • Researchers inform problem definition, ensuring chosen approach is technically feasible and ethically sound	• Provide strategic direction and establish desired functionalities for GenAI system	• Offer guidance on ethical considerations • Advise potential societal impacts of the proposed GenAI application	
Data Acquisition and Preprocessing	• Acquire data relevant to the problem definition • Prioritize data quality and address potential biases through careful selection and pre-processing techniques	• Provide legal expertise, advising on data privacy regulations and ensuring that data collection adheres to ethical and legal standards • Safeguard user privacy	• Offer best practices for data anonymization and responsible data handling • Contribute to ethical data management practices	

Model Development and Training	• Develop and train the GenAI model • Select algorithms and training data that minimize bias, mitigating potential discriminatory outputs	• Provide essential resources and oversight • Define how GenAI model fits into application or system		
Model Evaluation and Refinement	• Evaluate and refine the model's performance in terms of accuracy, fairness, and potential biases	• Align expectations and outcomes with model's performance in terms of business goals	• Analyze model outputs for ethical concerns and recommend mitigation strategies • Recommend metrics and evaluation frameworks for assessing bias and fairness in GenAI models	
Deployment		• Oversee deployment of GenAI system, ensuring it aligns with intended purpose and it delivers the desired functionalities • Guarantee compliance with relevant regulations on AI deployment and data usage		• Receive clear information on use of GenAI within the system they interact with • Report issues and concerns to ensure ongoing improvement
Monitoring and Maintenance	• Monitor model's performance to potentially address bias drift and safety risks that may emerge over time	• Allocate resources for ongoing maintenance	• Review and adapt regulations, keeping pace with other technological advancements	

Table 2. Strategies for effective stakeholder engagement across the GenAI development lifecycle

Transparent Communication Practices

Clear and honest communication about GenAI capabilities and limitations

Building public trust and fostering responsible adoption of Generative AI (GenAI) technologies necessitates communication approaches that prioritize transparency, user education, and balanced representation of both GenAI's strengths and limitations. Below are some methods in fostering informed public discourse and empowering stakeholders to make responsible decisions regarding GenAI development and deployment:

1. Leverage plain language: Technical jargon (e.g. Transformer, GANs) and complex terminology can be alienating for non-technical audiences. Communicate GenAI capabilities and limitations in clear, concise, and professional language tailored to the target audiences.

2. Highlight strengths and acknowledging weaknesses: Do not overpromise or make exaggerated claims about the GenAI system’s capabilities. Stakeholders developing the system should be upfront about its limitations in areas like reasoning, critical thinking, understanding complex contents, and handling ambiguity. Emphasize its strengths in automation, content creation, data analysis, and pattern recognition, while acknowledging that human expertise remains vital for tasks requiring judgment, creativity, and social understanding.

3. Emphasize transparency in outputs: Explain how outputs from GenAI systems are generated. This allows users to understand the inherent limitations and potential biases associated with AI-generated content.

4. Encourage user feedback mechanisms: Establish clear and accessible channels for users and stakeholders to provide feedback on GenAI outputs or interactions.

5. Proactive stakeholder education and awareness campaigns: Invest in educational resources and awareness campaigns about GenAI, not just to potential end-users, but to company employees and policymakers. This empowers different stakeholders to understand their potential responsibilities and the technology itself, while fostering informed and responsible use.

By implementing these strategies, stakeholders can cultivate a culture of open and honest communication around GenAI capabilities and limitations. This transparency engenders trust with users, empowers them to leverage the technology responsibly, and paves the way for a future where GenAI serves as a powerful tool for societal good while mitigating potential risks.

Reporting and disclosure practices

Strategic communication is two-way, and it can be through surveys, town hall sessions, interviews, and design thinking workshops. Although these methods may build the foundation of the stakeholder management, building the strategy afterwards shall but put together and aligned by identified participants who has the role of the sharing the strategic plan moving forward. Towards implementation, it is important to proactively follow the stakeholder map and engage the appropriate stakeholders. Finally, monitoring stakeholder feedback is necessary to be able to adjust and align the approach in the GenAI development process.

Regulatory and Legal Considerations

This section explores the current state of GenAI-related regulations across various jurisdictions and proposes action points on how stakeholders can proactively address compliance challenges.

Compliance with Existing Laws and Regulations

Overview of relevant laws

With the fast-phasing landscape of the use of AI technologies, different nations across the world are trying to catch up with the creation of laws to regulate and provide guidance of their uses to public to ensure that these promote greater good. One of the notable laws is the EU AI Act in which it classifies AI according to its risks. It adopts a risk-based approach for categorizing AI systems into four (4) tiers which generally corresponds to 1) sensitivity of the data involved and 2) the AI use case or application. The EU AI Act categorized the risk as 1) unacceptable risk, 2) high risk, 3) limited risk (AI systems with specific transparency obligations), and 4) minimal risk.

To highlight, AI practices that pose “unacceptable risk” are explicitly prohibited in the Act. This includes practices in marketing, providing or using AI-based systems that:

- Use manipulative, deceptive and/or subliminal techniques to influence a person to decide that they otherwise would not have made in a way that causes or may cause that person or other persons’ significant harm;
- Exploit vulnerabilities of person(s) due to their age, disability, and/or specific social/economic situation to influence the behavior of that person in a way that causes or may cause that person or other person's significant harm;

- Use biometric data to categorize individuals based on their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation; and
- Create or expand facial recognition databases through untargeted scraping of facial images from the internet or closed-circuit television (CCTV) footage (European Parliament, 2024).

In the United States, the AI regulation approach is decentralized and generally, most regulatory practices and policies are focused on sectoral levels. The US has established several sector-specific AI-related agencies and organizations that address some of the challenges arising from the evolution of AI. For instance, the Federal Trade Commission (FTC) addresses consumer protection by promoting fair and transparent business practices in AI applications (US Federal Trade Commission, 2024). Similarly, the National Highway Traffic Safety Administration (NHTSA) oversees the safety of AI-powered technologies, especially in autonomous vehicles (National Highway Traffic Safety Administration, 2023). Additionally, states like California have implemented their own regulations, such as the California Consumer Privacy Act (CCPA), which imposes stringent requirements on businesses processing consumer data, including those using AI (Department of Justice - State of California, 2018). Overall, while AI regulation in the United States is not centralized, it is supported by various sector-specific regulations.

On the other hand, China has emerged as a major player in the global AI sector, with ambitions to become the leading AI innovation center by 2030. The government is keenly aware of the ethical and security implications of AI, implementing measures to control its growth and operations. China's comprehensive AI and cybersecurity regulations reflect guiding principles aimed at ensuring data protection and risk management. Key frameworks, such as the Chinese Cybersecurity Law and the New Generation AI Development Plan, emphasize compliance and security. Balancing its drive for AI dominance with the need for ethical practices, China is aligning its strategies with global standards to solidify its position as a leading AI superpower (Filipova, 2024).

Canada has taken a proactive approach to AI regulation, balancing the promotion of innovation with the preservation of ethical standards and societal interests. The country has launched key initiatives like the Pan-Canadian AI Strategy and the Canadian AI Ethics Council to support responsible AI development and address ethical issues. Additionally, Canada’s Personal Information Protection and Electronic Documents Act (PIPEDA) regulates the collection, use, and disclosure of personal information in AI technologies, ensuring stringent data protection and the preservation of privacy rights (Department of Justice - Government of Canada, 2019).

We expect more nations to enforce new regulations as AI technology advances. AI regulations play a crucial role in shaping the future societal impact of technology. These

regulations should set clear standards and support AI across different sectors, prioritizing ethical and consumer protection principles. Collaboration among stakeholders at national and global levels is essential to ensure responsible AI implementation and to maximize its benefits.

Strategies for compliance

There is no "one-size fits all" in terms of compliance when we are dealing with AI regulations. Oftentimes, it requires a combination of best practices, various governance frameworks, sector-specific regulations and guidelines, and relevant territorial regulations. As an organization, one of the first actions is to check or revisit their internal policies, procedures (i.e. privacy policies, data use policies, information classification and management policies, terms of service) to ensure that these consider for planned or permitted uses of generative AI.

Ensuring compliance should also be looked at the enterprise level which requires collaboration among relevant internal stakeholders within an organization like enterprise risk management, procurement services, data privacy office, and information security. Through this internal collaboration, organizations may be able to determine whether transparency to the public or impacted individuals regarding the organization's generative AI use is necessary. Organizations operating in regulated industries must prioritize understanding and effectively communicating their specific legal obligations and liabilities. It is crucial for the organization to consult guidance provided by regulatory agencies regarding the use of generative AI, integrating this knowledge into their internal policies and procedures (Almond, 2023).

Proactively Shaping Future Regulations

Engaging with policymakers

Through proactive engagement with policymakers and stakeholders, organizations can contribute to shaping balanced and effective regulations that promote the responsible deployment and development of generative AI technologies. In the case of Aboitiz Data Innovation (ADI), we participate in various technical working groups in the government and provide inputs and insights to enable the implementation of the Philippines National AI Strategy Roadmap which is anchor in two (2) pillars: a.) Implementation that covers the strategic dimensions on digitalization and infrastructure, workforce development, and regulation; b.) Innovation that covers the strategic dimension on research and development (DOST-PCIEERD, 2021).

Another effort is we constantly collaborate with experts in various fields such as AI experts, legal experts, think tanks, academe, and industry leaders to develop well-informed policy

recommendations. Through this, we can present a unified front that demonstrates broad support and expertise in the field. In ADI, we also demonstrate responsible practices to showcase our commitment to responsible AI practices through use cases and best practices. Our initiatives highlight the steps taken to address ethical concerns and to comply with existing regulations.

Contributing to the development of new standards and guidelines

There are several steps that an organization can take to contribute to the development of new standards and guidelines which entails structured approach that requires research, collaboration, advocacy, and continuous engagement. In the recently published ASEAN Guide on AI Governance and Ethics, a document that serves as a practical guide for organizations in the ASEAN region to design, develop, and deploy traditional AI technologies in commercial and non-military or dual-use applications, ADI provided its own use case highlighting one of the key components --- Internal Governance Structures and Measures. Further, the guide mainly focuses on fostering the interoperability of AI frameworks across jurisdictions. This is one of the organization's ways to collaborate and contribute to the development of relevant policies and guidelines to foster robust, ethical, and effective standards for generative AI, ensuring that the technology benefits society while managing the risks (Association of Southeast Asian Nations, 2024).

Future Directions and Innovations in Generative AI Ethics

In this section, we investigate emerging areas of research and discussion that hold promise for mitigating ethical risks associated with AI, and how potential shifts in the regulatory landscape could affect the Framework and GenAI stakeholders.

Technological Advancements

The future of Generative AI lies in further advancements in deep learning techniques. As neural networks become more sophisticated, models can generate content with increased complexity, accuracy, and realism. Techniques like Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) will continue to evolve, yielding more diverse and compelling results (The Future of Generative AI: Trends and Predictions, 2024). Smaller, more efficient language models will also continue to improve, making AI accessible to a

wider audience. Open-source models will drive innovation and democratize AI development (Bergmann, 2024).

Generative AI will transcend domain-specific applications, embracing cross-domain creativity. Models that can process and generate content across different modalities—text, images, music, etc.—will become the norm. This will enable richer and more versatile applications, fostering innovation in various industries simultaneously. These models will also aid in content creation, idea generation, and even foster creative collaboration.

We must also consider the ethical implications of increasingly autonomous GenAI systems and the potential for adversarial uses of GenAI, such as deepfakes becoming so sophisticated that they are difficult to detect. Moreover, the spread of misinformation is particularly concerning due to the realistic presentation of the generated content, which can be frighteningly convincing and personalized to various personalities.

Emerging Trends in GenAI Ethics

As GenAI technology continues to advance, new ethical considerations emerge. In the following, we explore key aspects that need careful consideration. These are also captured in Figure 3.

- Data Responsibility and Privacy
    - Safeguarding User Data: As Generative AI systems evolve, safeguarding user data and privacy becomes paramount. Organizations must adopt robust data governance practices, ensuring transparency regarding data collection, usage, and retention.
    - Balancing Performance and Privacy: Striking a balance between model performance and privacy protection is essential. Organizations should be mindful of the ethical implications of using personal data for training AI models.

- Fairness, Transparency, and Inclusion
    - Mitigating Bias: Generative AI models must be designed to avoid discriminatory outcomes. Techniques for identifying and mitigating biases in training data and model outputs are critical.
    - Creating Fair and Representative Content: Addressing biases (in training data) is essential, ensuring that the generated content does not perpetuate stereotypes or reinforce existing inequalities. Looking ahead, Generative AI will increasingly emphasize the development of techniques to mitigate

biases, fostering the creation of fair, diverse, and representative content. Transparency mechanisms, such as Explainable AI (XAI), play a crucial role in helping users understand how decisions are made by these AI systems.

- Inclusion and Accessibility: Generative AI should benefit all users, regardless of their background or abilities. Ensuring accessibility means considering diverse user needs, including those with disabilities. Models should generate content that is inclusive, culturally sensitive, and relevant across different contexts.
- Ethical Metrics: Beyond traditional AI model metrics such as accuracy, fairness metrics, such as Disparate Impact and Equal Opportunity Difference, evaluate model performance form an ethical perspective and can be integrated into the AI model design process. These metrics measure the ratio of favorable outcomes for different groups.
- Environmental Sustainability: As AI adoption grows, minimizing its environmental impact is vital. Energy-efficient training methods, hardware optimizations, and responsible resource usage are essential. Organizations should explore ways to reduce the carbon footprint associated with AI development and deployment.

- Accountability, Trust, and Regulatory Landscape
  - Establishing Accountability Mechanisms: Organizations must take responsibility for the behavior of their AI systems. Regular audits, clear documentation, and adherence to ethical guidelines build trust with users. The usage of XAI can help build trust and identify potential biases within the system.
  - Human-in-the-Loop: Human oversight in GenAI workflows can mitigate risks and ensure responsible decision-making, especially for high-stakes applications.
  - Transparent Communication: When issues arise, transparent communication and corrective actions are essential.
  - Ownership and Misuse: As GenAI creates more realistic and impactful content (text, images, etc.), questions of ownership, attribution, and potential for misuse will need to be addressed.
  - Regulatory Influence: Organizations should anticipate evolving regulations and oversight bodies related to GenAI. The growing importance of user

education and awareness about GenAI capabilities and limitations cannot be overstated.

- User Education and Awareness: Educating users about GenAI capabilities and limitations is crucial in times where society evolves at a slower pace than the advancement of GenAI technologies.

Figure 3: Key aspects of GenAI Ethics

The Human-AI Relationship

As GenAI capabilities grow, fostering a collaborative and responsible human-AI relationship becomes crucial. Mechanisms for human oversight and control over GenAI systems will be essential to ensure responsible use. Exploring how humans and AI can work together effectively to leverage the strengths of each will be key to maximizing benefits and mitigating risks. The potential impact of GenAI on the workforce and the need for responsible job reskilling and upskilling initiatives must be considered. Furthermore, the importance of human values being embedded in the design and development of GenAI systems cannot be ignored.

The Role of Standardization

The question of how to approach GenAI ethics across different industries and regions is critical. Can we develop a standardized ethical framework for GenAI, or will frameworks need to be customized based on industry and cultural contexts? Fostering international collaboration on GenAI ethics is essential to ensure responsible development and deployment on a global scale. We will need to strike a balance between standardization and customization.

For example, the European Union (EU) is known for its comprehensive approach to data privacy with the General Data Protection Regulation (GDPR). This focus on privacy may influence how they approach standardization of GenAI ethics. The United States often takes a more industry-driven approach to regulations. This could lead to a preference for industry-specific best practices alongside broader ethical frameworks. China has its own unique cultural and legal system that will need to be considered when developing frameworks for GenAI ethics in that region.

Industry-specific best practices and guidelines can exist alongside broader ethical frameworks, and cultural and legal differences must be considered when developing these frameworks.

Long-Term Ethical Considerations and Global Perspectives

As GenAI capabilities continue to evolve, we face a range of profound ethical questions that extend beyond immediate applications and require a global perspective. Generative AI, which includes algorithms capable of creating text, images, audio, code, and videos, raises several ethical questions that distinguish it from traditional AI systems. For instance, generative AI can easily produce deepfakes, misleading or false information, and its outputs can appear trustworthy, leading to the spread of misinformation and the manipulation of public opinion. Additionally, there is a risk of generative AI inadvertently producing harmful or offensive content.

In the first set of ethical considerations below, GenAI presents a complex picture for human identity and society, varying across different cultures and regions and significantly impacting humans in both positive and negative ways.

- The Nature of Creativity and Originality: As GenAI becomes adept at generating creative outputs like art, music, and literature, it raises questions about the nature of human creativity and originality. Will AI-generated works be considered true art, and if so, who deserves credit: the programmer, the AI itself, or a combination?
- The Future of Work and Automation: The automation of tasks by GenAI is likely to continue, potentially displacing workers in various sectors. We need ongoing

discussions about the future of work, including the need for reskilling and upskilling initiatives to ensure a smooth transition for individuals whose jobs are impacted.

- Ethical Considerations for Social Justice and Equality: It is crucial to consider the potential societal implications of GenAI. Will it exacerbate existing inequalities, or can it be used to promote social justice? How can we ensure that the benefits of GenAI are distributed fairly across different demographics and socioeconomic groups?
- The Impact on Human Values and Decision-Making: As GenAI becomes more integrated into our lives, how will it influence our values and decision-making processes? We must ensure that AI systems are designed and developed in a way that aligns with human values and promotes responsible decision-making.

GenAI also has the potential to reshape global governance. While it offers opportunities for improved decision-making and collaboration, it also necessitates robust international cooperation to address ethical concerns and mitigate long-term risks.

- The Global Landscape of GenAI Ethics: Different countries have varying cultural and legal frameworks. The ethical considerations surrounding GenAI may differ based on societal values and existing regulations. International collaboration on GenAI ethics is essential to ensure responsible development and deployment on a global scale.
- The Potential for Existential Risks: Some experts raise concerns about the long-term potential for highly advanced AI to pose existential risks to humanity. Open discourse and careful planning are required to mitigate these potential risks.
- The Need for Continuous Discourse and Adaptation: The ethical landscape of GenAI is constantly evolving. It is essential to adapt ethical frameworks as technology advances and societal values change. This includes fostering a culture of responsible innovation and research, with regular assessments of potential risks and benefits.

By prioritizing these considerations, we can foster a future where GenAI wholly benefits humanity and contributes to a more just and equitable global society.

The Need for Ongoing Dialogue

Despite proliferation of AI policies and guidelines across countries, there is no universal consensus on ethical principles and standards for generative AI. Harmonizing these

frameworks is challenging and the lack of alignment hinders effective global cooperation (Kerry, Meltzer, Renda, Engler, & Fanni, 2021). Additionally, ethical norms differ globally. What is acceptable in one culture may not be acceptable in another culture, and bridging these gaps requires nuanced dialogues. Therefore, navigating the ethical landscape of GenAI requires continuous dialogue and adaptation.

Open communication and collaboration among developers, policymakers, researchers, and the public are crucial for developing robust ethical frameworks. These frameworks should be adaptable and flexible to accommodate future advancements and unforeseen challenges. The importance of ongoing research into the ethical implications of GenAI, and continuous education and awareness-raising initiatives to keep all stakeholders informed, cannot be overstated. By addressing these future directions and innovations in GenAI ethics, we can ensure that this powerful technology is developed and used responsibly for the benefit of all.

Conclusion

Generative AI (GenAI) offers a transformative force with immense potential to revolutionize numerous sectors. However, alongside its advantages lie significant ethical considerations that demand careful attention. This white paper has proposed a comprehensive GenAI Governance Framework, emphasizing the crucial role of transparency, explainability, fairness, and stakeholder engagement. We explored practical applications of this framework, highlighting key considerations for stakeholders involved in GenAI development, implementation, and regulation. Additionally, successful case studies demonstrated responsible GenAI integration within the banking industry.

All in all, we conclude that our proposed GenAI Governance Framework provides a critical roadmap for responsible GenAI development and deployment. Its lifecycle-based approach, domain-specific considerations, and clear distinction between its benefits and those of a framework-less approach makes it an essential tool for fostering a responsible and inevitable future with GenAI.

The Power of a Well-Defined Framework

Our proposed GenAI Governance Framework integrates eleven (11) key principles throughout the six (6) stages of the GenAI system development lifecycle. Considering these principles at each stage is crucial because it ensures ethical considerations are embedded throughout the entire process, from initial concept to ongoing monitoring. These principles

can be grouped into three categories, providing a roadmap for navigating GenAI development and deployment:

- Responsible Development and Use: This category focuses on ensuring GenAI is developed and used ethically, considering societal impacts and potential biases. It encompasses principles such as:
    - Ethical Use
    - Social Impact
    - Bias and Discrimination Mitigation
    - Innovation and Advancement
- Transparency and Accountability: This category emphasizes the importance of clear communication and establishing lines of responsibility throughout the GenAI lifecycle. It includes principles like:
    - Transparency and Explainability
    - Accountability and Responsibility
    - Data Protection and Privacy
    - Education and Training
- Safety and Ongoing Oversight: This category focuses on ensuring the safety and reliability of GenAI systems, including addressing potential risks. It includes the principles:
    - Safety and Reliability
    - Regulatory Compliance
    - Stakeholder Engagement (for ongoing feedback and adaptation)

By prioritizing these principles within a well-defined framework applied across each stage of the GenAI lifecycle – problem definition, data acquisition and preprocessing, model development and training, model evaluation and refinement, deployment, and monitoring and maintenance – we ensure a holistic approach where we address GenAI's ethical, technical, and societal aspects comprehensively. With this in place, we can envision a future where GenAI flourishes ethically, contributing to a more just, equitable, and prosperous world for all.

Call to Action: Fostering Ethical GenAI Development

Moving forward, continuous active collaboration among developers, policymakers, researchers, and the public is paramount. Open communication fosters trust and transparency, allowing for the enhancement of existing frameworks to be more robust and adaptive, accounting for future advancements and unforeseen challenges.

Continuous research plays a vital role in understanding the evolving ethical landscape of GenAI. As a result, organizations around the world will be more open to adapting these frameworks to their GenAI implementations, avoiding ethical failures like the Air Canada situation (Yagoda, 2024). Investing in research initiatives will provide valuable insights to guide ethical considerations. Furthermore, education and awareness-raising campaigns are crucial for keeping all stakeholders informed. By empowering all involved with a comprehensive understanding of GenAI's ethical implications, we can make informed decisions that prioritize responsible development and deployment.

References

Agnostelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., . . . Roberts, A. (2023, Jan 26). *MusicLM: Generating Music From Text.* Retrieved from Google Research: https://arxiv.org/pdf/2301.11325

Aldridge, A. (2024, 8 April). *BofA's Erica Surpasses 2 Billion Interactions, Helping 42 Million Clients Since Launch.* Retrieved from Bank of America - Newsroom: https://newsroom.bankofamerica.com/content/newsroom/press-releases/2024/04/bofa-s-erica-surpasses-2-billion-interactions--helping-42-millio.html

Almond, S. (2023, April 3). *Generative AI: eight questions that developers and users need to ask.* Retrieved from Information Commissioner's Office - United Kingdom: https://ico.org.uk/about-the-ico/media-centre/blog-generative-ai-eight-questions-that-developers-and-users-need-to-ask/

Amoozadeh, M., Daniels, D., Nam, D., Kumar, A., Chen, S., Hilton, M., . . . Alipour, M. A. (2024). Trust in Generative AI among Students: An exploratory study. *55th ACM Technical Symposium on Computer Science Education V. 1* (pp. 67-73). ACM.

Association of Southeast Asian Nations. (2024). *ASEAN Guide on AI Governance and Ethics.*

Avci, E. (2023). Generative AI in Electricity Distribution: A Qualitative Exploration. *PressAcademia Procedia*, 208-211.

Barde, K., & Kulkarni, P. (2023). Applications of Generative AI in Fintech. *Third International Conference on AI-ML Systems* (pp. 1-5). ACM.

Bergmann, D. (2024, February 9). *The most important AI trends in 2024.* Retrieved from IBM: https://www.ibm.com/blog/artificial-intelligence-trends/

Blakey, D. (2023, October 13). *Wells Fargo extends advice and planning, LifeSync access to all retail banking customers.* Retrieved from Retail Banker International: https://www.retailbankerinternational.com/news/wells-fargo-extends-life-sync-to-all-retail-banking-customers/

Brittain, B. (2024, April 20). *Bank of America defeats 'Erica' virtual-assistant trademark case on appeal.* Retrieved from Reuters: https://www.reuters.com/legal/litigation/bank-america-defeats-erica-virtual-assistant-trademark-case-appeal-2024-04-19/

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., . . . Shyam, P. (2020). Language models are few-shot learners. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, (pp. 1877-1901).

Buehler, K., Corsi, A., Weintraub, B., Jurisic, M., Siani, A., & Lerner, L. (2024, March 22). *Scaling gen AI in banking: Choosing the best operating model.* Retrieved from McKinsey & Company: https://www.mckinsey.com/industries/financial-services/our-insights/scaling-gen-ai-in-banking-choosing-the-best-operating-model#/

Cacciaglia, A. (2024, March 4). *Gen AI Saves Nurses Time by Drafting Responses to Patient Messages.* Retrieved from EpicShare: https://www.epicshare.org/share-and-learn/mayo-ai-message-responses

Chiu, C. (2024, May 29). *A brief history of ChatGPT.* Retrieved from Charity Digital: https://charitydigital.org.uk/topics/a-brief-history-of-chatgpt-11369

Clark, M., & Vincent, J. (2023, 15 March). *What's new with GPT-4 — from processing pictures to acing tests.* Retrieved from The Verge: https://www.theverge.com/2023/3/15/23640047/openai-gpt-4-differences-capabilties-functions

Department of Justice - Government of Canada. (2019). *Personal Information Protection and Electronic Documents Act.*

Department of Justice - State of California. (2018). *California Consumer Privacy Act.*

Doanh, D., Dufek, Z., Ejdys, J., Ginevicius, R., Korzynski, P., Mazurek, G., & Paliszkiewicz, J. (2023). Generative AI in the Manufacturing Process: Theoretical Considerations. *Engineering Management in Production and Services*, 76-89.

DOST-PCIEERD. (2021). *Artificial Intelligence and Information & Communications Technology - Roadmapping Executive Report.*

Dourado, R. (2021, June 5). *Is it ethical to replace human workers with automation?* Retrieved from Rossy Writes - Medium: https://rossywrites.medium.com/is-it-ethical-to-replace-human-workers-with-automation-6866ca1d5db

European Parliament. (2024, June 18). *EU AI Act: first regulation on artificial intelligence.* Retrieved from European Parliament: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Filipova, I. (2024). Legal Regulation of Artificial Intelligence: Experience of China. *Journal of Digital Technologies and Law*, 46-73.

Gaster, A. (2024, May 6). *Generative AI tools: Is Freepik new Reimagine any good?* Retrieved from Dorve: https://dorve.com/blog/ux-news-articles-archive/freepik-reimagine-ai-tool/

Germanidis, A. (2024, June 17). *Introducing Gen-3 Alpha.* Retrieved from Runway: https://runwayml.com/blog/introducing-gen-3-alpha/

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Proceedings of the 27th International Conference on Neural Information Processing Systems*, (pp. 2762-2680).

Hacker, P., Engel, A., & Mauer, M. (2023). Regulating ChatGPT and other Large Generative AI Models. *2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1112-1123). ACM.

Hanna, D. (2023). The Use of Artificial Intelligence Art Generator "Midjourney" in Artistic and Advertising Creativity. *Journal of Design Sciences and Applied Arts*, 42-58.

Hoover, A. (2024, April 9). *Students Are Likely Writing Millions of Papers With AI.* Retrieved from Wired: https://www.wired.com/story/student-papers-generative-ai-turnitin/

Hu, K. (2023, February 2). *ChatGPT sets record for fastest-growing user base - analyst note.* Retrieved from Reuters: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Kerry, C. F., Meltzer, J. P., Renda, A., Engler, A., & Fanni, R. (2021). *Strengthening international cooperation on AI.* Brookings.

Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. *International Conference on Learning Representations 2014.*

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*.

Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022). News from Generative Artificial Intelligence Is Believed Less. *2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 97-106). ACM.

Maastricht University. (2023, July 2). *ChatGPT Guidelines for Examiners.* Retrieved from Maastricht University: https://www.maastrichtuniversity.nl/file/22052-boe-fasos-chatgpt-guidelines-examiners-v-7-2-2023pdf

Maastrict University. (2024). *Large Language Models and Education.* Retrieved from Maastricht University: https://www.maastrichtuniversity.nl/education/edlab/ai-education-maastricht-university/large-language-models-and-education

Markowitz, D. M. (2024, April 23). *From Complexity to Clarity: How AI Enhances Perceptions of Scientists and the Public's Understanding of Science.* Retrieved from ArXiv: https://arxiv.org/abs/2405.00706

Marshall, M. (2024, January 12). *Wells Fargo's assistant, powered by Google's AI, poised to hit 100 million interactions annually.* Retrieved from VentureBeat: https://venturebeat.com/ai/wells-fargos-google-llm-driven-assistant-may-reach-100-million-interactions-per-year/

Martineau, K. (2023, April 20). *What is generative AI?* Retrieved from IBM Research: https://research.ibm.com/blog/what-is-generative-AI

Mesko, B., & Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *Nature*.

Miyazaki, K., Murayama, T., Uchiba, T., An, J., & Kwak, H. (2024). Public perception of generative AI on Twitter: an empirical study based on occupation and usage. *EPJ Data Science*.

National Highway Traffic Safety Administration. (2023). *Report to Congress - Automated Vehicles.*

National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework.* US Department of Commerce.

Oh, S., Kang, M., Moon, H., Choi, K., & Chon, B. S. (2023). A Demand-Driven Perspective on Generative Audio AI. *International Conference on Machine Learning.*

Papaj, N. (2022, October 24). *Wells Fargo's New Virtual Assistant, Fargo, to Be Powered by Google Cloud AI.* Retrieved from Wells Fargo - Newsroom: https://newsroom.wf.com/English/news-releases/news-release-details/2022/Wells-Fargos-New-Virtual-Assistant-Fargo-to-Be-Powered-by-Google-Cloud-AI/default.aspx

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial IntelligenceAssessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment. *Journal of University Teaching & Learning Practice*.

Personal Data Protection Commission Singapore, Infocomm Media Development Authority Singapore. (2020). *Model Artificial Intelligence Governance Framework.* Singapore.

Radwan, A., & McGinty, J. (2023). Toward a Conceptual Generative AI Ethical Framework in Teacher Education. In M. Searson, E. Langran, & J. Trumble, *Exploring New Horizons: Generative Artificial Intelligence and Teacher Education* (pp. 87-110). AACE - Association for the Advancement of Computing in Education.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., . . . Sutskever, I. (2021). Zero-Shot Text-to-Image Generation. *38th International Conference on Machine Learning* (pp. 8821-8831). PMLR.

Schlagwein, D., & Willcocks, L. (2023). 'ChatGPT et al.': The ethics of using (generative) artificial intelligence in research and science. *Journal of Information Technology*.

Seward, Z. (2024, May 10). *The New York Times introduces the A.I. Initiatives Team.* Retrieved from Editor & Publisher: https://www.editorandpublisher.com/stories/the-new-york-times-introduces-the-ai-initiatives-team,249706

Shekhar, D. (2023, May 22). *Google Cloud Text to Speech API: The Future of AI Voice Synthesis.* Retrieved from Medium: https://medium.com/@imdshekhar/google-cloud-text-to-speech-api-the-future-of-ai-voice-synthesis-a65db9ad688d

Takyar, A. (2024). *Generative AI in manufacturing: Use cases, benefits and development.* Retrieved from LewayHertz: https://www.leewayhertz.com/generative-ai-in-manufacturing/

The EU Artificial Intelligence Act. (2024). *The EU Artificial Intelligence Act*. Retrieved from The EU Artificial Intelligence Act: https://artificialintelligenceact.eu/

The Future of Generative AI: Trends and Predictions. (2024, March 23). Retrieved from Indika.ai: https://www.indikaai.com/blog/the-future-of-generative-ai

Thompson, A. D. (2024, February 9). *Synthesia.* Retrieved from Life Architect: https://lifearchitect.ai/synthesia/

US Federal Trade Commission. (2024). *Tech Summit on Artificial Intelligence: Consumer Facing Applications.*

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, (pp. 6000-6010).

Velasco, M. M. (2024, March 26). *Generative AI touted as next 'disruptor' in the energy sector.* Retrieved from Manila Bulletin: https://mb.com.ph/2024/3/25/generative-ai-touted-as-next-disruptor-in-the-energy-sector

Yagoda, M. (2024, February 23). *Airline held liable for its chatbot giving passenger bad advice - what this means for travellers.* Retrieved from BBC: https://www.bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know

Yin, S., Wu, C., Yang, H., Wang, J., Wang, X., Ni, M., . . . Yang, F. (2023). NUWA-XL: Diffusion over Diffusion for eXtremely Long Video Generation. *61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1309-1320). ACL.

Zlateva, P., Steshina, L., Petukhov, I., & Velev, D. (2024). A Conceptual Framework for Solving Ethical Issues in Generative Artificial Intelligence. *Electronics, COmmunications and Networks*, 110-119.

In addition, listed below are a curated set of references arranged by the different areas of literature surrounding GenAI governance:

Ethical Frameworks	
Towards a Conceptual General AI Ethical Framework in Teacher Education	https://www.researchgate.net/profile/Elizabeth-Langran-2/publication/379698349_Exploring_New_Horizons_Generative_Artificial_Intelligence_and_Teacher_Education_Published_by_AACE_-Association_for_the_Advancement_of_Computing_in_Education/links/661683ddf7d3fc28743fa531/Exploring-New-Horizons-Generative-Artificial-Intelligence-and-Teacher-Education-Published-by-AACE-Association-for-the-Advancement-of-Computing-in-Education.pdf#page=94
A Conceptual Framework for Solving Ethical Issues in Generative Artificial Intelligence	https://books.google.com.ph/books?hl=en&lr=&id=CYLyEAAAQBAJ&oi=fnd&pg=PA110&dq=generative+ai+ethical+frameworks&ots=lnljpZMd3X&sig=vzo2UqAzGjj8YsGcq3TtAsMWpkk&redir_esc=y#v=onepage&q=generative%20ai%20ethical%20frameworks&f=false
ChatGPT et al.' The ethics of using (generative) artificial intelligence in research and science	'ChatGPT et al.': The ethics of using (generative) artificial intelligence in research and science - Daniel Schlagwein, Leslie Willcocks, 2023 (sagepub.com)
The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment	https://open-publishing.org/journals/index.php/jutlp/article/view/810/769
ASEAN Guide on AI and Ethics v2	https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf
NIST AI Risk Management Framework	https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
Model AI Governance Framework	https://iapp.org/media/pdf/resource_center/pdpc_model_framework_ai_governance_second_edition.pdf
Regulatory Landscape	
From RAG to QA-RAG: Integrating Generative AI for Pharmaceutical Compliance Process	https://arxiv.org/pdf/2402.01717
The imperative for regulatory oversight of large language models (or Generative AI) in healthcare	https://www.nature.com/articles/s41746-023-00873-0
Regulating Generative AI: Ethical Considerations and Explainability Benchmarks	https://osf.io/h74gw/download/?format=pdf
Regulating ChatGPT and other Large Generative AI Models	https://arxiv.org/pdf/2302.02337

Generative AI and the AI Act	https://www.europarl.europa.eu/cmsdata/277774/02.YordankaIvanova.pdf
From Automation to Augmentation: Redefining Engineering Design and Manufacturing in the Age of Next-Gen AI	https://mit-genai.pubpub.org/pub/9s6690gd/release/2
On Generative Artificial Intelligence: Open-Source is the Way	https://osf.io/jnmzg/download
Public Perception and Trust	
Public perception of generative AI on Twitter: an empirical study based on occupation and usage	https://link.springer.com/content/pdf/10.1140/epids/s13688-023-00445-y
Generative Ominous Dataset: Testing the Current Public Perception of Generative Art	https://dl.acm.org/doi/fullHtml/10.1145/3623462.3623475
Trust in Generative AI among Students: An exploratory study	https://arxiv.org/pdf/2310.04631
Science Written by Generative AI is Perceived as Less Intelligent, but More Creditble and Trustworthy than Science Written by Humans	https://arxiv.org/pdf/2405.00706
News from Generative Artificial Intelligence in Believed Less	https://dl.acm.org/doi/fullHtml/10.1145/3531146.3533077
Bias and Fairness	
Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies	https://www.mdpi.com/2413-4155/6/1/3
Improving the Fairness of Deep Generative Models without Retraining	https://arxiv.org/pdf/2012.04842
Bias in Generative AI	https://arxiv.org/pdf/2403.02726
Fair Generative Modeling via Weak Supervision	https://proceedings.mlr.press/v119/choi20a/choi20a.pdf
Fair Generative Models via Transfer Learning	https://ojs.aaai.org/index.php/AAAI/article/view/25339/25111
On Measuring Fairness in Generative Models	https://proceedings.neurips.cc/paper_files/paper/2023/file/220165f9c7f51163b73c8c7fff578b4e-Paper-Conference.pdf
Achieving Causal Fairness through Generative Adversarial Networks	https://par.nsf.gov/servlets/purl/10126320
PreciseDebias: An Automatic Prompt Engineering Approach for Generative AI to Mitigate Image Demographic Biases	https://openaccess.thecvf.com/content/WACV2024/papers/Clemmer_PreciseDebias_An_Automatic_Prompt_Engineering_Approach_for_Generative_AI_To_WACV_2024_paper.pdf
Transparency and Explainability	
User Submission - Explainable Generative AI (XGenAI): Enhancing Transparency and Trust in AI Systems	https://indiaai.gov.in/article/explainable-generative-ai-xgenai-enhancing-transparency-and-trust-in-ai-systems
Investigating Explainability of Generative AI for Code through Scenario-based Design	https://dl.acm.org/doi/pdf/10.1145/3490099.3511119
Foregrounding Artist Opinions: A Survey Study on Transparency, Ownership, and Fairness in AI Generative Art	https://arxiv.org/pdf/2401.15497
The importance of transparency: Declaring the use of generative artificial intelligence (AI) in academic writing	https://sigmapubs.onlinelibrary.wiley.com/doi/full/10.1111/jnu.12938

Explainable Generative AI (GenXAI): A Survey, Conceptualization, and Research Agenda	https://arxiv.org/pdf/2404.09554
xAI-GEN: Enhancing Generative Adversarial Networks via Explainable AI Systems	https://arxiv.org/pdf/2002.10438
Towards Explainable NLP: A Generative Explanation Framework for Text Classification	https://arxiv.org/pdf/1811.00196
Privacy and Security	
Privacy and AI Governance Report	https://iapp.org/media/pdf/resource_center/privacy_ai_governance_report.pdf
Identifying and Mitigating the Security Risks of Generative AI	https://arxiv.org/pdf/2308.14840
Generative AI for Secure Physical Layer Communications: A Survey	https://arxiv.org/pdf/2402.13553
Generative AI for Cyber Security: Analyzing the Potential of ChatGPT, DALL-E, and Other Models for Enhancing the Security Space	https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10491270
Machine learning techniques for IoT security: Current research and future vision with generative AI and large language models	https://www.sciencedirect.com/science/article/pii/S2667345223000585
Context-Aware Generative Adversarial Privacy	https://www.mdpi.com/1099-4300/19/12/656
Privacy-Preserving High-dimensional Data Collection with Federated Generative Autoencoder	https://petsymposium.org/popets/2022/popets-2022-0024.pdf
Security and Privacy on Generative Data in AIGC: A Survey	https://arxiv.org/pdf/2309.09435
Privacy and Security Concerns in Generative AI: A Comprehensive Survey	https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10478883
Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges	https://www.jmir.org/2024/1/e53008/
Accountability and Responsibility	
Generative AI meets Responsible AI: Practical Challenges and Opportunities	https://sites.google.com/view/responsible-gen-ai-tutorial/
Editors' statement on the responsible use of generative AI technologies in scholarly journal publishing	https://link.springer.com/article/10.1007/s11019-023-10176-6
The Janus Effect of Generative AI: Charting the Path for Responsible Conduct of Scholarly Activities in Information Systems	https://pubsonline.informs.org/doi/epdf/10.1287/isre.2023.ed.v34.n2
Generative AI in Responsible Conversational Agent Integration: Guidelines for Service Managers	https://www.sciencedirect.com/science/article/pii/S0090261624000184
Protecting scientific integrity in an age of generative AI	https://www.pnas.org/doi/epdf/10.1073/pnas.2407886121
Responsible Generative AI: What to Generate and What Not	https://arxiv.org/pdf/2404.05783
Copyright Protection and Accountability of Generative AI: Attack, Watermarking and Attribution	https://arxiv.org/pdf/2303.09272
Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift	https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/718/399