Generative AI Legal, Privacy and Security Concerns

Ram N Sangwan

- **Concerns Around Legal, Privacy and Security**
- **Concerns Around IP**
- **Responsible AI**
- **Enterprise Best Practices**

Concerns Around Legal, Privacy and Security

Legal, Privacy and Security Concerns

Common Concerns

Intellectual Property Rights

GenAI can create content that may infringe on existing copyrights, trademarks, or patents, raising concerns about the ownership of AI-generated content.

Data Privacy

LLMs are trained on vast amounts of data, which could include personal or sensitive information, leading to privacy breaches if the data is not handled with strict confidentiality.

Consent and Image Rights

Using someone's likeness to generate new content without consent can violate their image rights, especially if the content is shared publicly or monetized.

Security Risks

AI systems can be exploited to generate deepfakes, contributing to the spread of misinformation, or to create malicious software, posing significant security threats.

Legal, Privacy and Security Concerns

Common Concerns

Bias and Discrimination

If AI is trained on biased data, it can perpetuate discrimination, leading to legal consequences under anti-discrimination laws.

Accountability and Liability

Determining who is responsible for the actions of AI systems (the creator, user, or the AI itself) is challenging, creating a complex legal landscape for liability.

Transparency and Explain ability

Many AI systems operate as "black boxes" with decision-making processes that are not transparent, which can lead to trust and accountability issues.

Regulatory Compliance

AI creators and users must navigate a patchwork of international laws and regulations, which can be difficult given the rapid advancement and global nature of AI technology.

Legal, Privacy and Security Concerns

Common Concerns

Access and Control

The question of who has the right to access and control AI systems and their outputs can lead to legal disputes, especially in competitive industries.

Ethical Use

There are ethical concerns about the use of generative AI, such as in creating fake content.

Informed Consent

For AI systems that interact with humans, obtaining informed consent for data collection and use is crucial.

Impact on Employment?

Legal, Privacy and Security Concerns

Common Concerns

Cybersecurity Measures

Organizations employing generative AI must implement robust cybersecurity measures to prevent unauthorized access and misuse of AI systems.

Cross-border Data Flow

AI systems often require the transfer of data across borders, which can conflict with national privacy laws and international agreements.

Digital Rights Management (DRM)

AI-generated content will challenge existing DRM systems, necessitating the development of new methods to protect and manage digital rights.

End-User Protection

Legal frameworks need to be developed to protect end-users from potential harm caused by interactions with AI, such as exposure to inappropriate content.

Concerns Around IP

Intellectual Property and Copyright Challenges

Generative AI tools operate without verifiable data governance and protection assurances.

This raises concerns about the exposure of confidential enterprise information.

Users should assume that synthetic data entered into generative AI platforms may become public, necessitating stringent controls to avoid inadvertent exposure of intellectual property (IP).

Companies must develop strategies to protect sensitive information and adhere to copyright laws.

Concerns around IP

Intellectual Property Concerns

Ownership of AI-Created Works

Who holds the copyright of works created by AI:

- The developer of the AI, the user who initiated the generation, or potentially no one.

Use of Copyrighted Material for Training

Generative AI may use copyrighted material during its learning phase, which raises questions about fair use and whether this constitutes infringement.

Digital Enforcement of IP Rights

The decentralized and often anonymous nature of AI-generated content dissemination makes it challenging for IP owners to monitor and enforce their rights.

https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html

Concerns around IP

Intellectual Property Concerns

Patentability of AI-Generated Inventions

As AI becomes capable of inventing, there's a debate about whether such inventions qualify for patents and, if so, who would be the inventor - the AI or the person who created the AI.

Trademark Infringement and Brand Damage

AI can generate content that includes trademarked terms or logos, potentially leading to brand dilution or confusion in the marketplace, and harming the brand's reputation and value.

Government Policy on AI

Ownership of content produced by AI tools

Indian copyright law does not currently recognize AI as an author, which leads to complexities regarding the ownership of AI-generated content.
This has been highlighted by the fact that the Indian Copyright Act does not explicitly include provisions for AI-generated content, which has led to debates and discussions on the need for amendments to address these new challenges
-- (India Today).

Government Policy on AI

Court Cases

In a recent case, the Delhi High Court held that using generative AI tools to portray famous personalities in fictional settings without their consent could violate their personal rights, which includes the right to privacy and publicity.
This shows that while AI can significantly contribute to various domains, its usage for commercial purposes without proper authorization can lead to legal infringements
-- (India Today).

Responsible AI

Responsible AI

Five points focusing on "Responsible AI" :

Ethical AI Design and Development

This involves creating AI with a focus on ethical considerations, such as fairness, accountability, and transparency, to ensure it operates within moral and societal norms.

Bias and Fairness

Responsible AI requires the identification and mitigation of biases in datasets and algorithms to prevent discrimination and ensure equitable outcomes for all users.

Transparency and Explainability

AI systems should be designed to be understandable by humans, with clear explanations for their decisions, to foster trust and accountability.

Privacy and Data Governance

Implementing robust data governance policies to protect user privacy and ensure that data is used responsibly in compliance with laws and ethical standards is critical for responsible AI.

Human-Centric AI

AI should augment human abilities and respect human dignity, ensuring that AI decisions are subject to human review, especially in critical applications such as healthcare or law enforcement.

An example

An example - Tata Consultancy Services (TCS)

TCS has been known to commit to ethical AI by establishing principles that govern their AI projects. These principles include transparency, explainability, fairness, auditability, and accountability.
TCS ensures that its AI systems are designed and deployed responsibly by adhering to these principles, which align with global standards and frameworks for responsible AI.
Moreover, TCS actively engages in AI research that focuses on ethical AI, bias detection, and the creation of AI systems that can explain their decisions to users, thus promoting trust and transparency.

Enterprise Best Practices

Responsible AI Principles – An Example

The excitement of Generative AI has never been higher.

- While this is very positive for anyone looking to build products and companies using Generative AI, it is also important we proceed responsibly.
- We are focusing on an Example for *“**Building our startup and our AI education product**. “*

Responsible AI Principles

We'll use the principles of Responsible AI:

- Fairness, Inclusiveness, Reliability/Safety, Security & Privacy, Transparency and Accountability.
- With these principles, we will explore how they relate to our use of Generative AI in our products.

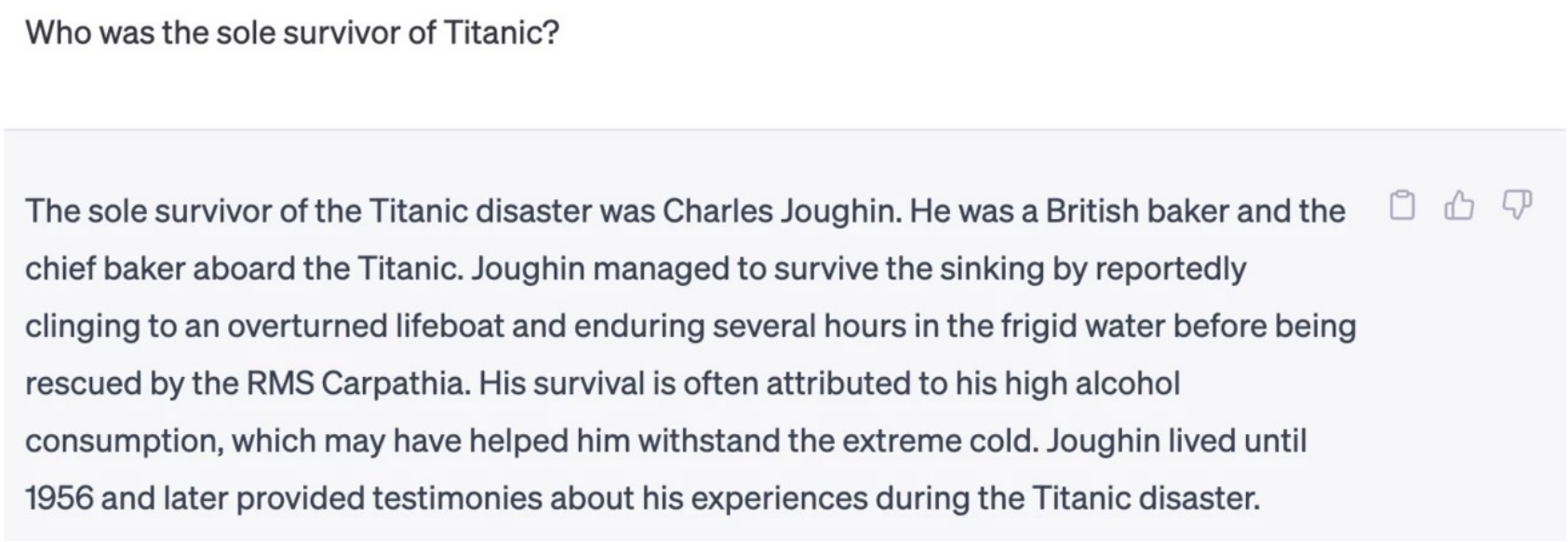
Why Should You Prioritize Responsible AI?

- When building a product, taking a human-centric approach by keeping your user's best interest in mind leads to the best results.
- The uniqueness of Generative AI is its power to create helpful answers, information, guidance, and content for users.
- This can be done without many manual steps which can lead to very impressive results.
- Without proper planning and strategies, it can also unfortunately lead to some harmful results for your users, your product, and society as a whole.

Hallucinations

A student asks the question ***Who was the sole survivor of Titanic?***

- Let's take for example we build a feature for our startup that allows students to ask historical questions to a model.



Hallucinations

… A student asks the question: ***Who was the sole survivor of Titanic?***

- This is a very confident and thorough answer.
- Unfortunately, it is incorrect.
- Even with a minimal amount of research, one would discover there was more than one survivor of the Titanic disaster.
- For a student who is just starting to research this topic, this answer can be persuasive enough to not be questioned and treated as fact. The consequences of this can lead to the AI system being unreliable and negatively impact the reputation of our startup.
- With each iteration of any given LLM, we have seen performance improvements around minimizing hallucinations. Even with this improvement, we as application builders and users still need to remain aware of these limitations.

Harmful Content

Harmful content can be defined as:

- Providing instructions or encouraging self-harm or harm to certain groups.
- Hateful or demeaning content.
- Guiding planning any type of attack or violent acts.
- Providing instructions on how to find illegal content or commit illegal acts.
- Displaying sexually explicit content.

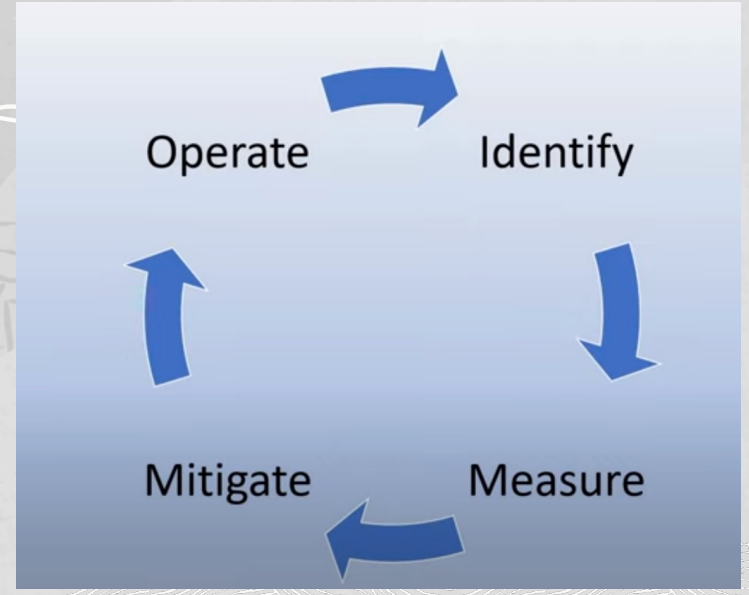
For our ***startup***, we want to make sure we have the right tools and strategies in place to prevent this type of content from being seen by students.

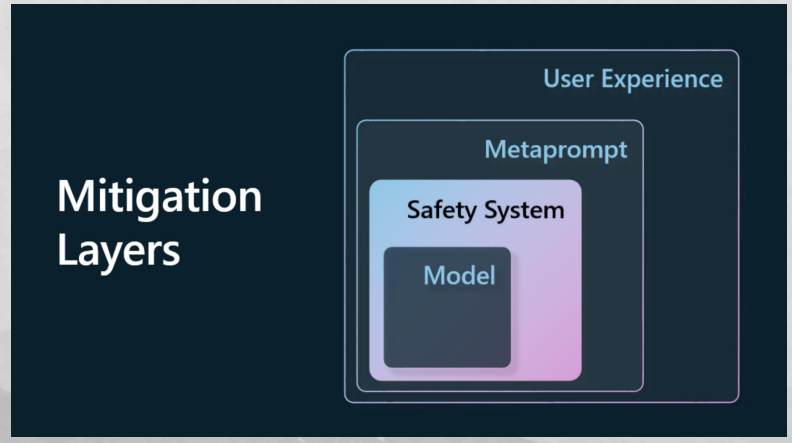
Lack of Fairness

Ensuring that an AI system is free from bias and discrimination and that they treat everyone fairly and equally.

- In the world of GenAI, we want to ensure that exclusionary worldviews of marginalized groups are not reinforced by the model's output.
- These types of outputs are not only destructive to building positive product experiences for our users, but they also cause further societal harm.
- As application builders, we should always keep a wide and diverse user base in mind when building solutions with Generative AI.

How to Use Generative AI Responsibly





Measure Potential Harms

- In software testing, we test the expected actions of a user on an application.
- Similarly, testing a diverse set of prompts users are most likely going to use is a good way to measure potential harm.
- Since our startup is building an education product, it would be good to prepare a list of education-related prompts.
- This could be to cover a certain subject, historical facts, and prompts about student life.

Choosing the right Model, Safety Systems and Prompts

Model, Safety Systems and Prompts

Model

- Choosing the right model for the right use case.
- Larger models can cause more of a risk when applied to smaller and more specific use cases.
- Using training data to fine-tune also reduces the risk of harmful content.

Safety System

- A set of tools and configurations on the platform serving the model that help mitigate harm.
- E.g., the content filtering system.
- Systems should also detect jailbreak attacks and unwanted activity like requests from bots

Metaprompt

- Metaprompts and grounding are ways we can direct or limit the model based on certain behaviours and information.
- This could be using system inputs to define certain limits of the model.
- Providing outputs that are more relevant to the scope or domain of the system.

Operate a Responsible Generative AI solution

Building an operational practice around your AI applications

- Building an operational practice around your AI applications is the final stage.
- This includes partnering with other parts of our start-up like Legal and Security to ensure we are compliant with all regulatory policies.
- Before launching, we also want to build plans around delivery, handling incidents, and rollback to prevent any harm to our users from growing.

Enterprise Best Practices

"Enterprise Best Practices" for the implementation and management of AI systems

Ethical Governance Framework

Establish a framework that includes policies, procedures, and oversight for AI deployments, ensuring alignment with business objectives and ethical standards.

Risk Management

Identify and assess potential risks associated with AI, such as operational, reputational, and ethical risks, and develop strategies to mitigate them.

Training and Awareness

Invest in training programs for employees to understand AI technologies, their capabilities, and their limitations, fostering a culture of informed and responsible AI use.

Data Management

Implement comprehensive data management practices, ensuring the quality, integrity, and security of the data used for training and operating AI systems.

Continuous Monitoring and Auditing

Regularly monitor and audit AI systems to ensure they function as intended, are secure from breaches or manipulations, and remain compliant with evolving regulations and ethical guidelines.

A real-world example of Enterprise Best Practices

In the context of AI - The operations of HDFC Bank.

- They have automated their Level 1 support with a well-trained AI chatbot named EVA, which has successfully handled millions of customer queries since its launch.
- This demonstrates the bank's commitment to leveraging AI for improving customer service while adhering to enterprise best practices.
- Another example is Happiest Minds Technologies, which embeds ethical principles into their platform through periodic code and data reviews.
- They focus on privacy and compliance, ensuring that sensitive features like race, gender, ethnicity, and religion are not used for decision-making, aligning with regulations such as the European Union's GDPR standards for AI governance.

Thank You