AI Governance Roundtable #1: Responsibility for AI

How should responsibility for AI systems be allocated between developer and deployer?

This is the first of a series of roundtables convened by AI Singapore for representatives from industry, government, and academia to discuss Responsible AI. Such discussions are typically too narrow and too broad. Too narrow in that a few voices dominate the discussion – notably those in the United States and Europe, with China sometimes included. Too broad in that discussion is often limited to generalities and principles. This project aims to address both aspects of this problem, involving a wider set of stakeholders — in particular those from Southeast Asia — in more focused discussions of specific challenges in the application of Responsible AI to particular questions.

Rapporteur: Norakmal Hakim Bin Norhashim, AI Singapore

Executive Summary

As AI systems become more complex and pervasive, properly allocating **responsibility** to developers, deployers, and other stakeholders will be an important aspect of **managing risk**.

Relevant **stakeholders** extend beyond developers and deployers to include regulators, users, and civil society actors including academia.

These diverse actors have distinct roles and interests. **Developers** may seek to limit liability for model usage and protect their intellectual property. **Deployers** look to innovate and create new markets while minimizing responsibility for factors beyond their control. **Users** need assurances of protection in their interactions with AI, while **governments**, as regulators, aim for balanced legislation that protects citizens without stifling innovation. **Academics** and other actors have a role to play in fundamental research (though model development itself is currently largely driven by the private sector) as well as debating ethics and the public policy around diverse use cases of AI.

Analogies can be made to other examples of risk management, such as comparing developers

to suppliers of raw materials or manufacturers of components in the context of product liability or comparing AI risk management to cloud platforms and data governance. These analogies highlight parallels between AI and existing technologies, as well as its unique aspects that may require new approaches.

There are several possible approaches. The **market** is attuned to the need for clarity, as exemplified by some developers offering indemnities for infringement of intellectual property rights by foundation models. **Contracts** are emerging as an important mechanism for sophisticated actors to agree on how to manage risk, especially between developers and deployers. However, concerns arise that the complexity of contractual language could particularly disadvantage general users. This acknowledgment of the potential complexity in contracts is underscored by the fact that much of data protection law utilizes the concept of notional or implied consent to protect users. **Insurance** is another way of managing the risk of liability for AI systems, though it operates most effectively when there is adequate information about the risks. In the face of many "unknown unknowns", **regulation** will be important to minimize or mitigate the risks to users in particular. "Under-regulation" is a concern, but so is "over-regulation" if it limits innovation or drives it elsewhere. For the near term, the focus is likely to be on sector-specific, risk-based regulations over broad, horizontal measures that target the underlying technologies.

Agility and flexibility will be required, with **ongoing monitoring and auditing** of AI systems to proactively manage risks as they emerge. This highlights the evolving nature of responsibilities, especially for developers in terms of system-monitoring and the potential need for mandatory audits.

These solutions collectively point towards a multifaceted approach to AI governance, involving various stakeholders and emphasizing the need for collaborative, balanced, and forward-thinking strategies.

Moving forward, a **collaborative and effective** approach to AI development and deployment will require measures across the AI lifecycle.

In the **development** phase, the emphasis should be on rigorous data governance standards and meaningful transparency about AI capabilities and limitations. The **deployment** phase calls for adherence to these standards and specific risk management protocols. The **use** phase stresses the importance of a shared responsibility framework, which while broadly applicable across all stages, takes on particular importance in this phase, alongside the recognition of end-user diversity. In parallel, **regulation** should embrace agile policymaking, maintaining balanced regulations, and harmonizing AI regulation across jurisdictions. Regulatory sandboxes at the research or invention stage may help align AI development with human values.

Introduction

As AI systems become more complex and pervasive, properly allocating responsibility to developers, deployers, and other stakeholders will be an important aspect of managing risk. To that end, the AI Governance pillar of AI Singapore convened a roundtable to delve into the critical issue of allocating responsibility and accountability within the AI ecosystem. The aim was to cultivate an environment where AI could be used responsibly and with clear lines of accountability. The forum brought together perspectives from industry, government, and academia. Key insights included the identification of new stakeholders who should bear what kind of responsibilities, like end users who supply prompts that initiate AI's outputs, and the examination of the dynamics of interactions *among* stakeholders like regulators, developers, and deployers. The discussion emphasized the intricate relationships among these groups and their influence on the AI ecosystem.

A significant focus of the discussion was the current challenges in the AI ecosystem, notably the lack of clarity in stakeholder interactions and the establishment of sound operating principles within the ecosystem itself. One of the prominent issues raised was the ambiguity in allocating responsibility in the event of AI failures. To navigate these complexities, participants frequently drew parallels between AI and existing technologies, using analogies to ground the discussion in familiar contexts, such as product liability.

Additionally, the roundtable was a platform for exploring various solutions to these emerging issues. Solutions were examined from multiple perspectives, including market-driven approaches like contracts and insurance models, regulatory measures, and sector-driven protocols, such as transparency guidelines. The discussions delved into the benefits and potential challenges of each approach, reflecting a comprehensive evaluation of potential strategies to address the challenges in responsibility allocation in the AI ecosystem. These dialogues provided nuanced insights into how different solutions could be implemented effectively while considering their possible impacts on the diverse stakeholders within the AI landscape.

This report encapsulates the key points and insights from the roundtable discussions. It is structured into three main sections: outlining the identified responsibility allocation problems within the AI ecosystem, exploring potential solutions, and proposing the next steps. Each section aims to distil the collective wisdom and perspectives shared during the roundtable, offering a comprehensive overview of the current state and future directions of Responsible AI development and deployment.

1 Understanding the Problem

Though the initial question posed was the relative responsibility of developers and deployers, there are also other stakeholders that need to be taken into account. This section identifies those stakeholders, highlighting their interests and some of the issues they may confront.

1.1 Stakeholders

Developers are recognized as the creators of AI technology, holding the reins of tech IP. They aim to market their models while avoiding significant liability for deployer actions and seeking to protect their intellectual property and trade secrets. Yet, they face central roles in data governance debates and the ethical sourcing of development data. Pressure mounts for them to be transparent about their models' limitations and potential deployment challenges.

Deployers, as innovators utilizing developer-created AI models, interface directly with end-users and have tools that can significantly influence AI outcomes. They seek to manage risks and avoid liability for uncontrollable factors, with some entities that seek stability and transparency in the AI landscape favouring regulation, and others preferring rapid, disruptive, and unimpeded development. There are also entities that fall somewhere in between. Key challenges for them are shaping regulatory policies, managing data governance, and necessitating robust communication with developers for successful AI governance.

Users of AI, typically interacting with AI products implemented on deployers' platforms, are bound by deployer-set terms and need assurances of protection in their AI interactions. They are encouraged to accept some responsibility in AI usage, especially in relation to their inputs. A key issue is the psychological distance from AI control mechanisms, leading to excessive trust and a diminished sense of responsibility.

Governments act as regulators, drafting legislation and policies impacting AI. They strive for a regulatory balance that neither stifles innovation nor exposes citizens to undue risks. A significant challenge is keeping up with rapidly advancing technology to develop comprehensive and agile regulations.

Academia, the original inventors of algorithms and models driving AI advancements, emphasize the need for humanities and social sciences input to fully grasp AI's impact on human decision-making. They face a lack of oversight in early development stages, advocating for more stringent testing during the invention phase and clear criteria for model development and publication.

In sum, the AI development and deployment ecosystem is a complex network of intertwined responsibilities and interests among various stakeholders, each facing distinct challenges that must be addressed for responsible and ethical AI progression.

1.2 Analogies

Who is best placed to manage the risks in AI development? The use of analogies plays a pivotal role in anchoring our understanding of risk assessment within the AI lifecycle. These analogies enabled participants to draw parallels with familiar scenarios and operational environments, thereby providing a clearer perspective on the complex nature of AI risks.

- **AI developers as suppliers of raw materials vs. manufacturers of components:** This analogy delves into whether developers in the AI ecosystem are comparable to suppliers of fundamental elements, such as raw materials or operating systems, or if they more closely resemble makers of specific components, creating specialized parts for the broader AI product produced by deployers.

  The distinction hinges upon whether there are significant design inputs by the developers that result in increased variability in the final AI product's functions, thereby significantly elevating the developer's liability. This contrasts with viewing AI as raw materials, where AI models are commodities with minimal differentiation in their design or function.

  Furthermore, if indeed there are significant design inputs, then it is likely that only the developers have detailed knowledge about the AI's characteristics or flaws. In contrast, if a normative understanding of such information exists, or if this information is accessible, as seen with open-source models, this could in turn reduce the developers' liability. Having access to detailed information about the inner workings of AI, whether through open or controlled proprietary channels, is essential for appropriately allocating accountability.

  By drawing parallels with existing product liability paradigms, this perspective raises a crucial question: if developers are treated as component makers, could consumers trace accountability up the production line in case of issues? Conversely, if developers' roles are seen as fundamental and universal, akin to providing raw materials or operating systems, their liability for specific AI use cases might be more limited, focusing on the general integrity of the technology rather than its particular applications.

- **AI risk management through the lens of related technology:** This perspective looks at whether we can draw analogies between AI and relevant existing technology, such as cloud platforms. In this view, stakeholders can draw on existing frameworks that advocate a shared responsibility model, emphasizing collaborative involvement in managing AI risks.

- **AI risk management through the lens of data governance:** In this view, stakeholders managing the training of models have control over the data input, thereby shouldering risks associated with outputs that could be linked to the training data. Such risks include issues emerging from generative AI, like infringing on the IP of artists, or potential biases

within AI models. While developers traditionally carry the most responsibility, the growing capability of deployers to fine-tune models by adding their own training data tailored to their needs indicates a need for a more equitable distribution of responsibility between developers and deployers.

- **Developing AI in the way institutions train graduates:** This analogy likens AI developers to educational institutions. Much like how institutions provide education but are not responsible for their graduates' actions, developers, especially in the realm of open source, often do not bear liability for how their technology is deployed and used. This similarity is particularly pronounced in open-source releases, which commonly operate under licenses that limit developers' liability.

1.3 Issues Related to Risk Management

In their attempts at mitigating risks, participants have noted that stakeholders may face the following issues:

- **Unclear AI regulations and risks:** Deployers, as the entity frequently directly regulated, are often the party confronted with the challenge of meeting regulatory demands and often require detailed information from developers. This need is exacerbated by the current regulatory landscape's ambiguity in specifying which party is responsible for certain obligations. Without clear guidance, deployers might compel developers to provide more information than is necessary or required. This situation arises even in instances where deployers may not be the most appropriate party to fulfil these obligations, highlighting the inefficiencies in current AI regulations.

- **Intellectual property vs. transparency:** Developers may face a dilemma when finding the right balance between protecting their intellectual property and providing transparency in their models. (Transparency and explainability are the focus of AI Governance Roundtable #2.)

- **Diminished responsibility and overreliance on AI:** Users' psychological distance from AI systems can lead to a diminished sense of responsibility and an overreliance on AI. This issue is critical in areas like autonomous vehicles and weapons systems, where it can impact ethical reasoning and decision-making by users.

1.4 Open Source

The role of open source in AI development presents unique challenges. Traditional regulations that focus on gatekeepers like big tech companies are complicated by the diffusion of responsibility in open-source projects. This is particularly so for open-source AI models that enable developers to further finetune and alter model outputs for specific needs. Such projects can include numerous independent contributors, including potentially rogue actors.

Regulations tend to target larger entities, as seen with frameworks like GDPR, but this approach may be less effective for open-source AI since it often involves smaller-scale developers. Despite this gap in oversight, open-source AI carries risks similar to those developed by larger companies, underscoring the need for compliance and congruity in regulation.

2 Possible Solutions

Throughout the discussions, the exploration of potential solutions to the identified challenges in the AI ecosystem has been a focal point. These dialogues frequently involved a diverse range of stakeholders, underscoring the intricate interplay and the multifaceted nature of the AI ecosystem. The solutions covered range from market-driven approaches with a light regulatory touch to calls for entirely new regulatory frameworks. Participants often leaned towards solutions that followed the path of least resistance, utilizing analogies to demonstrate how existing operational frameworks could serve as an initial step towards addressing AI challenges. However, there was also recognition of the limitations inherent in current methodologies, with some participants advocating for the need to develop a new paradigm tailored to the unique demands of AI. This section delves into the complex interplay between solutions discussed in our roundtable, their potential shortcomings, and the evolving perspectives on how they might need to adapt to more effectively meet the emerging needs of the AI landscape.

2.1 Market

The market is already demonstrating movement towards addressing AI responsibilities, particularly as it is good for businesses to be seen as effectively implementing Responsible AI practices. For instance, some model developers are offering indemnities to deployers and users for IP violations in the development process. In particular sectors like finance, the responsibility placed on banks is prompting them to negotiate specific terms with developers, especially since banks are held directly liable for issues like failures and biases in AI systems. However, this evolving market approach carries the risk of market failure, where the natural market mechanisms may not sufficiently address all concerns related to AI development and deployment. For example, in the early development stages, AI developers may not have enough economic incentives to implement Responsible AI practices, or the market could consist of insufficiently differentiated options, preventing deployers from selecting based on Responsible AI standards.

2.2 Contract

The use of contracts to define limitations and responsibilities in AI development and deployment is becoming increasingly relevant. Contracts can stipulate deployer usage terms, such as using the product “as is,” but there are limits to how far risk can be contractually outsourced, considering laws like the Unfair Contract Terms Act (UCTA). There is a discussion about whether responsibility limitations should be based on degrees of control, questioning if developers should dictate how their products should or should not be used.

Contracts, seen as complementary to governmental regulation, may become necessary to clearly define and manage risk and responsibility between developers and deployers, detailing obligations, liabilities, and expectations. Yet, challenges arise in scenarios where a "take it or leave it" approach prevails due to limited developer options, stemming from a lack of distinct choices among developers based on their commitment to Responsible AI practices, and whether standard "terms and conditions" are sufficient for users to clearly understand the responsibility allocation between stakeholders (i.e. developers, deployers, users).

2.3 Insurance

Insurance as a solution to AI-related risks offers a safety net for users and deployers against potential liabilities. This model could serve as part of a broader strategy for addressing accidents or malfunctions in AI systems, providing financial coverage for damages. The analogy with vehicles indicates that, while insurance can distribute risk, it might not cover all scenarios, potentially leaving gaps in protection. There is also concern that reliance on insurance solutions might not fully address the complexities of the AI ecosystem, especially regarding accountability and ethical decision-making. Additionally, the presence of insurance might lead to a diminished emphasis on maintaining high ethical standards or thorough risk management practices.

2.4 Regulation

The need for "Goldilocks" regulation — not too restrictive or lenient — is highlighted, along with the importance of agile government regulation. The debate is whether governments should continue regulating outcomes, as seen with banks, or shift towards more direct regulation of the technological processes by developers and deployers, like the EU AI Act. A sectoral approach grounded in a risk-based framework is recommended, prioritizing targeted regulation over broad, horizontal measures that regulate AI technology as a whole. However, concerns arise that regulators here and abroad may lack experience and understanding of AI, which could complicate effective regulation. Additionally, the necessity for agility in regulation could create confusion if rules keep changing frequently.

2.5 Transparency

Beyond allocating post-facto responsibility after incidents occur, ongoing supervision through monitoring, audits, and recalls is suggested. Questions arise about whether developers should have continuous obligations to monitor system usage and whether system audits should be mandatory. In the event of systemic failures, developers are often best positioned to implement systemic changes. This approach emphasizes the need for transparency throughout the entire AI lifecycle.

Each of these solutions offers a pathway to manage the complexities of AI development and deployment, with its own set of advantages and challenges. Balancing these solutions effectively is key to achieving a responsible and ethical AI landscape.

3 Next Steps

As we progress towards a more integrated and coherent approach to AI development and deployment, the "Next Steps" section aims to harmonize perspectives across the diverse array of stakeholders in the AI ecosystem. It operates under the premise that, although specific stakeholders are often seen as "owning" particular stages of the AI lifecycle, such as developers with the creation of AI models, and deployers with their implementation, the responsibility for Responsible AI development and deployment is a shared endeavour.

These recommendations are designed to foster collaborative efforts and ensure that the AI ecosystem evolves in a beneficial, ethical, and effective manner. Concluding with a summary table, this section provides a clear and concise overview of the key actions and responsibilities, offering a roadmap for stakeholders to collectively advance in ensuring Responsible AI.

3.1 Research or Invention

In the research or invention phase, the focus should be on creating environments conducive to innovation while aligning AI development with human values.

- **Creating regulatory sandboxes for broad human-AI alignment:** Introducing sandbox environments at the research or invention stage can help identify fundamental risks early and ensure comprehensive alignment between human values and AI advancements. This approach has been trialled in the past in specific sectors (finance, for example) and may be possible in some cross-sector use cases to encourage the balanced and ethical development of AI technologies.

3.2 Development

In the development phase, the focus is on establishing high standards and clear protocols to guide the responsible creation of AI technologies.

- **Upholding rigorous data governance standards:** Developers, having substantial control over the training data, thereby directly influencing the resulting outputs of AI, must adopt sector-wide data governance standards. These standards should prioritize safeguarding data privacy, respecting the intellectual property of training data, and actively working to prevent biases. This adherence to high standards is vital for ensuring responsible and trustworthy AI outputs.

- **Promoting transparency in AI capabilities and limitations:** Given the risk of inadequate incentivization for transparency resulting in limited choices of developers with Responsible AI practices in the market, it is crucial for developers to maintain an industry culture of transparency about what their AI models can and cannot do. Clear communication about the strengths and limitations of AI models will provide deployers with essential insights, aiding in informed decision-making.

- **Establishing protocols for AI service disruptions:** It is imperative to develop protocols that outline the extent of developmental intellectual property to be shared with deployers and regulators. This transparency is key in building trust, especially in scenarios of failure. Clear government regulations on disclosure and stakeholder notification in the case of AI service disruptions are also essential.

3.3 Deployment

The deployment phase involves both adhering to established standards and developing specific risk management protocols.

- **Ensuring shared data governance standards:** Deployers, influencing AI outputs through methods like finetuning, must maintain adherence to data governance standards. This continuous commitment ensures the responsible use of AI throughout its deployment stage.

- **Focal point for AI risk management and responsibility allocation:** Deployers, who are in direct contact with users and are deeply integrated into their sectors, have a unique perspective on the risks of applying AI in their contexts. Their expertise could be leveraged to develop specific risk management protocols at this stage. Managing AI application risks involves identifying responsibility gaps between developers and deployers, which can be managed through contractual agreements. Deployers are also well-placed to conduct thorough testing of AI models to ensure compliance with sector-specific standards.

3.4 Usage

The usage phase calls for a shared responsibility framework and acknowledgment of user diversity in AI applications.

- **Adopting a shared responsibility framework:** While a regulatory framework supporting shared responsibility is applicable across all phases, it becomes crucial in the use phase, where users range from organizations to individuals and from sophisticated players to general users. Thus, in this phase, establishing clear regulations to delineate the distribution of responsibilities for ensuring user safety and managing the consequences of user-initiated prompts is imperative. Deployers should also actively instruct users on their responsibilities when utilizing AI technologies.

- **Acknowledging user diversity:** Deployers have the opportunity to tailor AI deployment to accommodate diverse user requirements. This approach is crucial as AI is a transformative technology impacting a wide range of users.

3.5 Regulations

Regulatory efforts should focus on supporting agile policymaking, maintaining balanced regulations, and harmonizing standards.

- **Promoting the development of a shared responsibility framework adapted for each phase:** Given that entities can simultaneously assume multiple roles as developers, deployers, and users, establishing shared responsibility frameworks on a phase-by-phase basis is necessary. This approach ensures that responsibilities are clearly defined and adapted to the unique challenges and dynamics of each phase. Regulators, in this context, can play a pivotal role by facilitating dialogue and cooperation among all parties to promote the development and adoption of these frameworks.

- **Supporting industries impacted by agile policymaking:** The rapid evolution of AI necessitates an agile approach to policymaking. Regulators should assist slower-moving industries in adapting to these changes.

- **Maintaining "Goldilocks" regulations:** Regulations should be carefully crafted to avoid overcomplication that might stifle AI innovation. Striking the right balance is key to fostering a productive AI ecosystem.

- **Harmonizing AI regulations across jurisdictions:** Governments can play a pivotal role in harmonizing AI regulations and standards, keeping costs manageable for developers and deployers and facilitating smoother cross-jurisdictional operations.

Each of these next steps is designed to guide stakeholders through the complex landscape of AI development and deployment, ensuring that AI technologies are developed, deployed, and used in a responsible, ethical, and effective manner.

Refer to the table at the end of the report for a summary of the steps and the role each stakeholder can play in ensuring a Responsible AI ecosystem.

4 Conclusion

The roundtable hosted by AI Singapore's AI Governance pillar highlighted the effectiveness of a grounded, stakeholder-centric approach in navigating the realm of Responsible AI. By focusing on tangible issues encountered by stakeholders, rather than abstract ethical concepts, the discussion pinpointed practical challenges and solutions within the AI ecosystem.

The use of analogies to relate AI to existing operating contexts was instrumental in the discussions, making complex AI problems more comprehensible. This approach allowed experienced stakeholders to better understand and address these challenges within known frameworks. However, alongside this, the roundtable also highlighted the necessity for new framings and paradigms in comprehending AI's unique aspects.

The diversity of solutions explored — encompassing market-driven strategies like insurance and contracts, alongside regulatory measures, and sector guidelines — illustrates the complexity of problem-solving in AI. This breadth of approaches reinforces the need for stakeholders' collaboration, showing how varied perspectives can synergize to create more holistic and effective strategies.

Ultimately, these rich, layered discussions revealed that the path to Responsible AI is paved with collaborative efforts across the AI ecosystem. Engaging all stakeholders in the AI lifecycle ensures a broad range of insights, leading to an AI landscape that is not only responsible and effective but also attuned to societal needs. This collective whole-of-ecosystem accountability and shared responsibility is essential as we journey into a world of ever-growing AI capabilities.

Funding: This roundtable was funded via a charitable grant from Google.org, as part of Google's Digital Futures Project.

Responsibility Allocation in AI Lifecycle Toward Responsible AI Ecosystems				
Actors/Stages	**Invention/Research**	**Development**	**Deployment**	**Usage**
Developers	Participate in regulatory sandboxes for human-AI alignment	(1) Uphold data governance standards (2) Promote transparency in AI functionalities (3) Establish protocols for AI service disruptions	(1) Uphold data governance standards during model finetuning (2) Participate in risk management (a) Identify responsibility gaps between developers and deployers	Tailor AI for user diversity
Deployers		Establish protocols on AI service disruptions	(1) Uphold data governance standards (2) Develop a sector-specific AI risk management framework (a) Identify responsibility gaps between developers and deployers (b) Perform testing of AI models to ensure compliance with sector-specific regulatory standards	(1) Tailor AI for user diversity (2) Educate users on responsibilities of using AI (3) Validate AI models in accordance with regulations
Users				Understand role and adhere to safe and Responsible AI use
Academia	Participate in regulatory sandboxes for human-AI alignment			
Regulators	Create regulatory sandbox for human-AI alignment	Establish protocols on AI service disruptions		Define shared responsibility framework for user safety
	(1) Promote the creation and adoption of shared responsibility frameworks across different phases (2) Agile policy making (3) Avoid overcomplicating regulations (4) Harmonize AI regulations across jurisdictions			