How Explainable Must Enterprise AI Be? Transparency, Accountability and Governance in Practice

By Eddy So and Betty Liang

Article 1 of this series explored the difference between retaining rights in and control over data, and retaining the value generated from that data. Article 2 examined how the value, control and accountability of Enterprise AI are allocated. This article turns to explainability, considering what it means in practice.

Organisations are deploying Enterprise AI into customer-facing, operational and decision-support processes. Boards, management teams and regulators increasingly ask whether AI systems are explainable. Yet Enterprise AI often relies on technologies that organisations do not own, cannot inspect and may never fully understand due to the "black box" phenomenon. The "black box" problem describes modern AI systems whose internal workings are not practically inspectable by and are opaque to their own developers or deployers. The AI system produces an output, but no one can point to a single line of reasoning in the way a person could explain a manual decision.

An organisation may not be able to inspect the underlying model or fully understand how the AI system arrived at a particular point, yet it may still need to explain decisions and investigate incidents in order to satisfy regulators' enquiries and maintain trust among stakeholders and the public.

Enterprise AI often creates a "visibility gap": organisations may have limited visibility into how a system operates while remaining fully accountable for how it is used.

The real challenge is, therefore, not whether every aspect of an AI system can be explained, but whether the organisation has sufficient visibility to govern the technology responsibly and remain accountable for its use.

Explainability Means Different Things to Different People

For present purposes, explainability can be understood as the ability of a system, or the organisation deploying it, to provide a clear and coherent account of a specific prediction or decision after the fact, even if the underlying model itself is not fully interpretable. Different stakeholders may ask different questions when it comes to explainability of an AI system.

- For boards and senior management, the focus is usually governance and accountability. The key question is straightforward: can we justify relying on this system?
- For General Counsel and compliance teams, explainability is often an accountability question. If concerns arise, can the organisation obtain the information needed to explain what happened, investigate the issue and demonstrate appropriate oversight?
- Procurement teams approach the issue differently. Their concern is what information, commitments and contractual rights the supplier will provide.
- Regulators typically view explainability through the lens of accountability: can the organisation demonstrate effective governance, monitoring and human oversight?

> **"The question is not whether an organisation fully understands how an AI model works. The question is whether it has sufficient visibility to remain accountable for how the system is used."**

What You Can and Cannot Get from an AI Vendor

Many organisations respond by asking AI vendors for more information about how systems reach their decisions. That is a reasonable instinct, but it does not resolve the underlying issue. In practice, there are limits to what can realistically be obtained, regardless of who is asking or how the request is framed.

Organisations can often access system documentation, model cards, governance policies, risk disclosures, security certifications and performance information. These materials may provide valuable insight into how a system should be deployed and governed.

However, some information is rarely available or is even non-existent due to the "black box" effect. Customers generally cannot obtain complete training datasets, proprietary source code or detailed internal development processes. Commercial confidentiality, intellectual property protection and supply-chain complexity make full transparency unrealistic for most Enterprise AI deployments.

The practical question is therefore not what information exists, but what information an organisation is entitled to obtain and rely on when accountability requires answers.

Explainability Is Really About Accountability

An organisation may never fully understand the foundation AI model it deploys. Nevertheless, it may still need to explain decisions influenced or created by that model, investigate complaints, justify governance choices, respond to regulators and answer questions from customers and stakeholders.

The objective is not perfect explainability but sufficient accountability. In practice, this means the organisation has enough visibility to understand relevant risks, oversee deployment, monitor performance, investigate incidents and support governance decisions when needed. The organisation should therefore be able to answer the following questions when there is an unexpected output:

- What happened?
- Why did it happen?
- What controls and oversight existed at the time?

Explainability therefore becomes less of a technical challenge and more of a governance question. The issue is not whether every element of the model can be interpreted. It is whether the organisation possesses sufficient information and oversight to remain accountable for outcomes. Boards rarely need to understand AI model architecture. They do need to understand whether appropriate governance, oversight and escalation mechanisms exist. Perfect explainability may be unrealistic. Responsible accountability is not.

How Explainability Becomes a Governance and Contracting Issue

From a governance perspective, organisations increasingly need oversight structures, escalation processes, monitoring arrangements and accountability frameworks that allow them to supervise AI-enabled activities.

The Hong Kong Government's Digital Policy Office originally built its Ethical Artificial Intelligence Framework for use within government departments, and it is now used as a reference point for any organisation planning, designing and implementing AI systems in Hong Kong. The framework rests on 12 ethical AI principles, including transparency and explainability, human oversight, fairness, accountability and lawfulness. For its governance structure, it recommends three tiers of defence: a project team handling day-to-day design and early risk mitigation; a project steering committee and assurance team providing independent review and approval; and the IT board or chief information officer, supported by external advisers where appropriate, responsible for ongoing monitoring of AI strategy and risk.

Other Hong Kong regulators point in the same direction. The Privacy Commissioner for Personal Data's Artificial Intelligence: Model Personal Data Protection Framework focuses on accountability, risk assessment, human oversight, transparency and governance across the full AI lifecycle. The Hong Kong Monetary Authority's ("HKMA") guidance on generative AI in customer-facing applications builds on its earlier principles for AI applications by requiring authorised institutions to maintain "human-in-the-loop", "human-on-the-loop" or "human-in-command" arrangements and to maintain clear governance and accountability for generative AI outcomes. HKMA recently published the Supervisory Guidance for Responsible Adoption of Artificial Intelligence and introduced the "ART" Framework to support the responsible adoption of AI, where one of the focuses is accountability throughout the AI lifecycle. The above instruments should be distinguished from legislation, but their practical significance is substantial, each is likely to shape what regulators regard as a reasonable standard for responsible AI governance and what governance measures are expected to be in place.

In Mainland China, binding rules translate transparency and accountability into specific obligations. The Interim Measures for the Management of Generative Artificial Intelligence Services ("Interim Measures") apply to generative AI services offered to the public in Mainland China, but exclude enterprise research, development and applications that do not provide such services to the public. Within that scope, the Interim Measures require effective steps to enhance service transparency and improve the accuracy and reliability of generated content, disclosure of intended user groups and mechanisms for handling complaints. Importantly, providers must cooperate with regulatory inspections, explain matters including training data sources and algorithmic mechanisms as requested, and provide necessary technical and data support. Although not obliged to record every piece of information used, organisations providing generative AI services to the public in Mainland China are expected to maintain the information and compliance arrangements needed to respond to regulatory inspections, while observing applicable record-preservation duties and restrictions on retaining user inputs and usage records.

The European Union's AI Act illustrates this in detail. It sorts AI systems into four tiers of risk: (1) "unacceptable risk" systems are prohibited outright; (2) "high-risk" systems face the EU AI Act's most demanding obligations, which include a risk management system, data governance controls, technical documentation, record-keeping (logging), human oversight and accuracy, robustness and cybersecurity requirements; (3) "limited-risk" systems carry narrower duties, such as informing the end-users that they are interacting with AI; and (4) "minimal-risk" systems face no specific obligations under the AI Act.

The United States illustrates a different, more fragmented regulatory picture. At the federal level, the National Institute of Standards and Technology's AI Risk Management Framework is voluntary rather than mandatory, but it has become a widely used reference point. It organises AI governance into four functions:

(1) Govern – developing a risk management framework within the organisation; (2) Map – establishing the context to frame risks related to an AI system; (3) Measure – employing different tools, techniques and methodologies to analyse, assess, benchmark and monitor AI risk and related impacts; and (4) Manage – allocating resources to mapped and measured risks regularly and having plans to respond to, recover from and communicate about incidents. Similar themes are also emerging at state level in the United States. For example, the Colorado Automated Decision-Making Technology Act will take effect on 1 January 2027 with requirements relating to disclosure on intended uses, known limitations and training data, human review and accountability for AI-assisted decisions, etc.

The regulatory developments around the world differ in detail but point in the same direction. Organisations are expected to maintain enough oversight and documentation to account for how an AI system is used rather than to explain how the underlying model works. Meeting that expectation in practice also depends on the contractual arrangements between the AI system developers and deployers.

Rather than seeking impossible levels of technical transparency, organisations should consider whether they have appropriate rights to obtain the information needed to govern and oversee the system responsibly. Five questions matter most:

- what information must the AI provider disclose?
- what information can be independently verified?
- can inputs, outputs and decisions be reproduced if an issue arises?
- will the supplier assist with investigations, audits and regulatory enquiries?
- can important outputs be reviewed, challenged or overridden?

Incident and investigation support may be particularly important. When concerns arise about AI-assisted decisions, organisations are often expected to explain what happened, assess the impact and demonstrate appropriate oversight. Those obligations may arise even where the underlying technology is supplied by a third party. Organisations should therefore consider whether contractual arrangements require suppliers to provide timely cooperation, relevant documentation, records, technical assistance and support for investigations, audits and regulatory enquiries.

These five questions determine whether an organisation can remain accountable in practice. What many organisations describe as "explainability" is often a combination of information rights, oversight mechanisms and governance controls that provide sufficient visibility to understand, investigate and account for the use of AI systems.

Conclusion

Organisations often ask whether an AI system is explainable. That is not always the most useful question. Enterprise AI will frequently involve technologies that cannot be fully inspected or understood by those deploying them, yet accountability does not disappear simply because visibility is limited.

The more practical question is whether an organisation has sufficient transparency, oversight and information to govern the AI responsibly. In many Enterprise AI deployments, accountability depends not on complete visibility into the model itself, but on whether organisations have access to the information needed to explain, investigate and govern how the technology is used. Organisations may not need complete visibility into how an AI model works. They do, however, need enough visibility to remain accountable for how it is used.

How We Can Help

As organisations continue to adopt Enterprise AI, governance expectations are evolving rapidly across jurisdictions. We assist clients with AI governance and accountability reviews, Enterprise AI procurement and vendor assessments, contract review and negotiation of transparency-related provisions, Board and management briefings on AI governance obligations, and AI risk management and oversight frameworks.

The appropriate level of explainability depends on the organisation's objectives, risk profile and regulatory environment. The key challenge is not perfect transparency, but ensuring accountability remains possible even when complete explainability is not.

About the Authors

Eddy So is a Partner at MinterEllison Hong Kong. He advises organisations on governance, regulatory and risk management issues, including accountability, oversight and technology risk arising from the adoption of new technologies.

Betty Liang is a Trainee Solicitor at MinterEllison Hong Kong. She is inspired to explore technology regulation, data governance, the risks involved in new technology adoption and the available risk management strategies.



Eddy So
Partner, Hong Kong
T: 2841 6879
E: eddy.so@minterellison.com



Betty Liang
Trainee Solicitor, Hong Kong
T: 2841 6951
E: betty.liang@minterellison.com

MinterEllison Group and Associated Offices:
Adelaide Auckland Brisbane Canberra Darwin Gold Coast Hong Kong London Melbourne Perth Sydney Wellington

This bulletin was prepared by MinterEllison LLP for highlighting certain issues at the relevant time and for information only, and is not intended to be comprehensive. It is not for providing legal advice, and its contents do not constitute legal advice and shall not be relied upon by anyone for any general purpose or in relation to specific transactions and/or circumstances. Professional advice should be sought before applying the information to particular circumstances. Please contact one of your regular contacts at MinterEllison LLP should you have any questions or comments on this bulletin. MinterEllison LLP accepts no responsibility for any loss which may arise from reliance on the information contained in this bulletin.

© 2026 MinterEllison LLP