The "third way" chosen by OpenAI

Tasuku Itoh

10 September 2025

Executive Summary

AI developers' governance models' societal impact has started to garner attention in the wake of generative AI's rapidly growing influence. In particular, OpenAI's conversion of its commercial subsidiary into a public benefit corporation (PBC) presents new institutional possibilities for pursuit of both profits and the public interest. Going forward, evaluation of AI developers' governance models may become a key consideration in users' AI selection decisions.

Tasuku Itoh

Principal Consultant

Financial AI Platform Promotion Department

Why OpenAI opted against fully for-profit model

In May 2025, OpenAI announced a major change to its business structure. Instead of moving toward full commercialization as previously planned, the company decided to turn its commercial subsidiary into a Delaware Public Benefit Corporation (PBC). The key takeaway from its decision is that the transition to a PBC will institutionally ensure the commercial subsidiary is more accountable and better aligned with the public interest. Meanwhile, OpenAI will maintain its existing two-tier governance structure with its commercial operations controlled by an NPO as the NPO's subsidiary.

With OpenAI recently valued at an estimated $500bn, its decision was driven by structural factors and external pressure. First, OpenAI reportedly needs to raise fresh capital to the tune of at least tens of billions of dollars[1]. It requires vast computational resources to operate ChatGPT, which has 800mn weekly users. Amid globally intensifying competition to develop AI technologies, OpenAI had no choice but to adopt an organizational structure that affords more flexibility in terms of raising capital.

Second, OpenAI has been facing mounting societal pressure, including legal threats and strong criticism from various quarters. Most notably, it was sued by its co-founder Elon Musk for allegedly betraying its founding principles, accused by AI researchers and citizen groups of sacrificing safety in pursuit of profits and targeted by petitions asking California and Delaware's Attorneys General to block its for-profit conversion.

OpenAI's decision to become a PBC was a compromise after facing criticism over

NOTE

1) In late March 2025, Softbank agreed invest up to an additional $40bn in OpenAI (to be reduced to $20bn if OpenAI does not complete its then-pending for-profit conversion).

its push toward a for-profit model. By keeping a nonprofit business model, while converting its commercial arm into a non-capped-profit PBC (which functions much like a regular for-profit company), OpenAI can stay true to its original mission and still provide stronger incentives for investors.

PBCs as third option for AI developers

The PBC is a relatively new form of corporate entity. It originated in the state of Maryland in 2010 and has since been adopted by at least 34 other US states to date. PBCs are now widely recognized as a hybrid between a for-profit company and an NPO.

PBCs' most salient characteristic is that they are duty-bound by their corporate charter to not only maximize shareholder value but also pursue the public interest. In a regular joint-stock company, management's primary responsibility is maximizing shareholder value. In a PBC, the Board of Directors has a legal responsibility when making decisions to consider certain public-interest objectives stipulated in the articles of incorporation. Examples of such objectives include environmental preservation, health promotion and advancement of cultural, artistic or scientific aims. Another duty required of PBCs is accountability to society. Prominent PBCs include US crowdfunding platform Kickstarter, OpenAI's rival Anthropic and outdoor goods retailer Patagonia.

2) A hallucination is plausible but inaccurate generative AI output.

The AI industry's interest in PBCs stems from risks posed by the unfettered pursuit of profits. One such risk is that generative AI models' output could be influenced by their owners' interests just as search engines and social media platforms' algorithms are tweaked to benefit advertising revenues. Another risk is social bias (i.e., the risk of AI output being biased in favor of or against certain demographics), which can be a form of generative AI hallucination[2], together with factually inaccurate output. A third concern is that AI output pertaining to geopolitical issues may be skewed by the views of the developer or its owners, as exemplified by the Chinese LLM DeepSeek.

Comparison of major AI companies' governance models

3) AI red teaming is organized hacking with the intent of detecting AI systems' vulnerabilities and risks. Red teaming originated in the cyber security space before being applied to AI safety assessments.

Major AI developers are currently exploring different approaches to building optimal governance regimes (see table). Google DeepMind is committed to strengthening AI safeguards through such means as AI red-teaming[3], establishing

an AGI Safety Council and annually publishing Responsible AI Progress Reports while driving leading-edge AI R&D with its vast financial resources and technological capabilities. However, it has been criticized for not unequivocally opposing militarization of AI. Some observers question whether it prioritizes societal responsibility on a par with profit maximization.

Anthropic has built an independent oversight regime by establishing not only a PBC but also a long-term benefit trust (LTBT)[4] that appoints its directors. Its Responsible Scaling Policy[5] mandates that its development program be suspended if the AI risk level exceeds a certain threshold.

4) The LTBT is an independent oversight entity established by but separate from Anthropic. It oversees Anthropic's decision-making from the standpoint of the long-term public interest.

5) Anthropic formulated its Responsible Scaling Policy (RSP) as a framework for managing risks of AI development. The RSP defines a hierarchy of risk levels for assessing AI systems' safety and is the AI industry's first explicit safety management policy that mandates cessation of development if certain risk thresholds are exceeded.

Meta is a purely for-profit company but it endeavors to ensure transparency through open-sourcing. However, such openness poses concerns about bad-actor risk.

In sum, AI developers are all seeking an optimal governance model in their own way as AI's societal influence grows. While no perfect solution exists yet, the emergence of PBCs and other such organizational frameworks as a viable option is a noteworthy development.

AI companies' governance models and safeguards

Company	Governance model	Safeguards
OpenAI	NPO + For-profit (PBC)	• SOC 2 Type II attestation • Preparedness team
Google DeepMind	For-profit	• AI red-teaming • AGI Safety Council • Responsibility and Safety Council • Responsible AI Progress Reports
Anthropic	PBC + LTBT	• ISO/IEC 42001 certification • Collaborations with US AI Safety Institute, METR etc. • Disclosure of AI model safety assessments • Responsible Scaling Policy
Meta	For-profit	• Transparency through open-sourcing • AI red-teaming • Model release testing

Source: NRI, based on company disclosures

Governance as new AI selection criterion

In the wake of AI's rapid development, major AI developers are experimenting with a variety of governance models, including solo PBC, NPO + PBC, PBC + trust, and fully for-profit. In Japan, most AI developers are still for-profit entities, none of which has shown any signs of pursuing the public interest in parallel with profits. Many Japanese companies are currently focused on formulating operational rules

for using AI. Few if any have reached the point of comprehensively addressing how to use AI or how to evaluate AI developers' governance.

When selecting AI technologies, companies tend to prioritize functionality, usability and/or cost. Going forward, it will become increasingly important to make such decisions based on a broader set of criteria, including AI developers' governance models and commitment to societal accountability.

about NRI

Founded in 1965, Nomura Research Institute (NRI) is a leading global provider of system solutions and consulting services with annual sales above $5.1 billion. NRI offers clients holistic support of all aspects of operations from back- to front-office, with NRI's research expertise and innovative solutions as well as understanding of operational challenges faced by financial services firms. The clients include broker-dealers, asset managers, banks and insurance providers. NRI has its offices globally including Tokyo, New York, London, Beijing and Sydney, with over 16,700 employees.

For more information, visit https://www.nri.com/en