By: Christopher Marquis

Tech leaders believe that ethics is optional. Americans must prove them wrong

As AI advances and increasingly penetrates our lives, it is unlikely either to create a technological utopia or to wipe out humanity.

The more probable outcome is somewhere in the middle – a future shaped by contingency, compromise, and, crucially, the decisions we make now about how to constrain and guide AI's development.

As the global leader in AI, the United States plays an especially important role in shaping that future. But US President Donald Trump's recently announced AI Action Plan has dashed hopes of strengthened federal oversight, embracing instead a pro-growth approach to developing the technology.

That makes it even more urgent for state governments, investors, and the American public to focus on a less-discussed tool for accountability: corporate governance.

As journalist Karen Hao documents in her book Empire of AI, the industry's leading firms are already engaging in mass surveillance, exploiting their workers, and exacerbating climate change.

The irony is that many are public-benefit corporations (PBCs), a governance structure purportedly designed to avoid such abuses and protect humanity. Clearly, it is not working as intended.

The structuring of AI companies as PBCs has been a wildly successful form of ethics-washing. By virtue-signaling to regulators and the public, these firms create a veneer of accountability that allows them to avoid more systemic oversight of their day-to-day practices, which remain opaque and potentially harmful.

For example, Elon Musk's xAI is a PBC whose stated mission is to "understand the universe."

But the company's actions – from furtively constructing a polluting supercomputer near a predominantly Black neighborhood in Memphis, Tennessee, to creating a chatbot that praises Hitler – demonstrate a troubling indifference to transparency, ethical oversight, and affected communities.

Lofty ambitions

PBCs are a promising tool for enabling companies to serve the public good while also pursuing profit.

But in its current form, the model – especially under the law of Delaware, the state where most US public companies are domiciled – is riddled with loopholes and weak enforcement, and thus cannot provide guardrails for AI development.

To prevent perverse outcomes, improve oversight, and ensure that firms incorporate the public interest in their operating principles, state legislators, investors, and the public must demand that PBCs are reimagined and strengthened.

Companies cannot be evaluated or held accountable without specific, time-bound, and quantifiable goals.

Consider how PBCs in the AI sector rely on sweeping, undefined benefit statements that allegedly guide operations.

OpenAI proclaims that their goal is to "ensure AGI benefits all of humanity," while Anthropic aims to "maximize positive outcomes for humanity in the long run."

> Lofty ambitions are meant to inspire, but their vagueness can be used to justify almost any course of action – including ones that jeopardize public welfare

These lofty ambitions are meant to inspire, but their vagueness can be used to justify almost any course of action – including ones that jeopardize public welfare.

But Delaware law does not require companies to operationalize their public benefit through

measurable standards or independent assessments.

And while it requires biennial reporting on benefit performance, the findings do not have to be made public. Companies can fulfill – or neglect – their obligations behind closed doors, with the broader public none the wiser.

As for enforcement, shareholders can theoretically sue if they believe the board has failed to uphold the company's public-benefit mission.

But this is a hollow remedy, because the harms from AI are typically diffuse, long-term, and external to shareholders.

The affected stakeholders – such as marginalized communities and underpaid contractors – have no practical avenues for recourse.

More than a reputational shield

To play a meaningful role in AI governance, the PBC model must act as more than a reputational shield.

That means changing how "public benefit" is defined, governed, measured, and protected over time. Given the lack of federal oversight, reforming this structure must be done at the state level.

PBCs should be forced to commit to specific, measurable, and time-bound objectives that are written into their governing documents, backed by internal policies, and tied to performance reviews, bonuses, and career advancement.

Clearly defined objectives, not vague aspirations, will help firms create the foundation for credible internal alignment and external accountability

For an AI firm, these goals could include

ensuring the safety of foundation models, reducing bias in model outputs, minimizing the carbon footprint of training and deployment cycles, implementing fair labor practices, and training engineers and product managers on human rights, ethics, and participatory design.

Clearly defined objectives, not vague aspirations, will help firms create the foundation for credible internal alignment and external accountability.

Governing boards and the oversight process must also be reimagined. Boards should include directors with verifiable expertise in AI ethics, safety, social impact, and sustainability.

Each firm should have a chief ethics officer

Each firm should have a chief ethics officer with a clear mandate, independent authority, and direct access to the board.

These officers should oversee ethical review processes and be given the authority to halt or reshape product plans when necessary.

Trump's AI Action Plan has confirmed his administration's unwillingness to regulate this fast-moving sector

Lastly, AI companies that are structured as PBCs should be required to publish detailed annual reports that include granular, disaggregated data related to safety and security, bias and fairness, social and environmental impact, and data governance.

Independent audits – conducted by experts in AI, ethics, environmental science, and labor rights – should assess the validity of this data, as well as the firm's governance practices and overall alignment with public-benefit goals.

Trump's AI Action Plan has confirmed his administration's unwillingness to regulate this fast-moving sector.

But even in the absence of federal oversight, state legislators, investors, and the public can strengthen corporate AI governance by pushing for reforms to the PBC model.

More and more tech leaders seem to believe that ethics is optional.

Americans must prove them wrong, or else let disinformation, inequality, labor abuse, and unchecked corporate power shape the future of AI.

Christopher Marquis is Professor of Management at the University of Cambridge.