# HOW TRUMP CAN MAKE AI SAFE - INSIDE SOURCES

*Posted on January 18, 2025 by Judd Rosenblatt | INSIDE SOURCES*

The United States is racing to build the most powerful technology in human history — and the truth is that nobody knows how to keep it under human control. The stakes could not be higher for President-elect Donald Trump, and the imperative is clear: a Manhattan Project can ensure we build AI systems that reliably serve U.S. interests and values while maintaining our technological edge.

The numbers tell a clear story. In 2024, $200 billion was spent on AI model training, equivalent to nearly a quarter of America's military budget. OpenAI employees are projecting AI could replace most intellectual labor within 18 months.

Meanwhile, Jacob Helberg — newly appointed Under Secretary of State for Economic Growth and a leading voice on the congressional commission on Chinese competition — has called for a Manhattan Project-style program to win the AI race.

The commission is right about the urgency. In our rush to maintain American supremacy in artificial intelligence, we're overlooking a fundamental challenge in determining whether AI ushers in unprecedented prosperity or catastrophe: ensuring these increasingly powerful systems remain fundamentally "aligned" with American values and interests.

This isn't sci-fi speculation. AI is advancing along an exponential curve that humans — naturally thinking linearly — consistently underestimate. Even the godfathers of modern AI, Yoshua Bengio and Geoffrey Hinton, are now sounding the alarm, reflecting a pattern. The closer experts are to cutting-edge AI development, the more they grasp this exponential trajectory and fear its risks.

There are serious concerns about the risks that AI poses, but there is an opportunity for the next president to get it right.

Trump's return to office presents a historic opportunity to reshape America's AI strategy to ensure we not only win the AI race but win it safely — building systems that remain reliably under American control and aligned with American interests.

Unaligned AI would threaten American sovereignty just as much as a Chinese-built AI. By solving alignment first, America can achieve lasting technological dominance rather than merely winning a preliminary sprint toward an uncontrollable technology.

Conservatives, in the tradition of Herman Kahn, have always excelled at "thinking the unthinkable." During the Cold War, Kahn forced policymakers to confront the genuine possibility of thermonuclear war. We must now apply that same clear-eyed approach to AI. We don't know if or when AI will pose an existential threat, but America's future hangs in the balance. We must not allow shortsightedness — or dispersed bureaucratic oversight — to jeopardize it.

The solution must match the exponential nature of the emerging challenge: a Manhattan Project-style initiative focused on massively increased funding for neglected, moonshot approaches to AI alignment. This strategy would simultaneously pursue multiple

unconventional research directions, maximizing breakthrough chances while maintaining America's technological leadership.

Like the maverick scientists behind history's most significant breakthroughs, solving alignment requires venturing beyond established paradigms and challenging fundamental assumptions. While early private-sector efforts show promise, scattered individual efforts aren't enough — to solve alignment, this research strategy needs rapid scaling across the public and private sectors.

A Manhattan Project for AI alignment would amplify these efforts, and the evidence increasingly suggests that the best alignment research enhances capabilities precisely because it makes systems fundamentally safer and more reliable.

America's adversaries recognize the stakes. Of course, while it may be naive to take the Chinese at face value, there is some evidence China's president, Xi Jinping, is an "AI doomer." Xi has acknowledged that AI will determine "the fate of all mankind" and must remain controllable. President Biden and Xi's November meeting in Peru saw the Chinese leader frame AI as a shared global challenge, ostensibly calling for expanded international dialogue and cooperation.

Luckily, China has not made the massive investments in computing infrastructure needed to compete in advanced AI development. Its strategy focuses on replicating Western advances rather than pursuing independent breakthroughsThe United States needs to engage in aggressive foreign policy that allows us to continue making AI progress while also avoiding moves that would force China into viewing AI development as a zero-sum race.

This strategic moment calls for Trump to flex his deal-making skills. His leadership on complex international agreements like the Abraham Accords proves he can rise to such moments. By combining his "America First" approach with a clear-eyed view of the global AI landscape, Trump could forge a responsible AI development framework that prioritizes alignment research while maintaining American technological supremacy.

By pursuing multiple ambitious research directions simultaneously through a Manhattan Project-style initiative, the administration can maximize our chances of breakthrough solutions while ensuring America maintains its technological edge. This is how we ensure AI development ushers in unprecedented American prosperity and freedom — by making systems that are not just powerful but fundamentally safe and aligned with our values.

The commission is right that we need a Manhattan Project for AI. We must ensure that the project prioritizes capabilities and AI alignment. Only then can we secure America's technological future, ensuring a world where our children survive and flourish.