



# AI Scientists Gather To Plot Solutions To Doomsday Scenarios

Technocrats have created a monster that they now realize could turn and destroy their creators; thus, 'war games' are created to learn how to defeat it - Artificial Intelligence. This is another example of an attempt to 'save the world' has turned into a 'destroy the world' scenario. □ TN Editor

Artificial intelligence boosters predict a brave new world of flying cars and cancer cures. Detractors worry about a future where humans are enslaved to an evil race of robot overlords. Veteran AI scientist Eric Horvitz and Doomsday Clock guru Lawrence Krauss, seeking a middle ground, gathered a group of experts in the Arizona desert to discuss the worst that could possibly happen — and how to stop it.

Their workshop took place last weekend at Arizona State University with

funding from Tesla Inc. co-founder Elon Musk and Skype co-founder Jaan Tallinn. Officially dubbed “Envisioning and Addressing Adverse AI Outcomes,” it was a kind of AI doomsday games that organized some 40 scientists, cyber-security experts and policy wonks into groups of attackers — the red team — and defenders — blue team — playing out AI-gone-very-wrong scenarios, ranging from stock-market manipulation to global warfare.

Horvitz is optimistic — a good thing because machine intelligence is his life’s work — but some other, more dystopian-minded backers of the project seemed to find his outlook too positive when plans for this event started about two years ago, said Krauss, a theoretical physicist who directs ASU’s Origins Project, the program running the workshop. Yet Horvitz said that for these technologies to move forward successfully and to earn broad public confidence, all concerns must be fully aired and addressed.

“There is huge potential for AI to transform so many aspects of our society in so many ways. At the same time, there are rough edges and potential downsides, like any technology,” said Horvitz, managing director of Microsoft’s Research Lab in Redmond, Washington. “To maximally gain from the upside we also have to think through possible outcomes in more detail than we have before and think about how we’d deal with them.”

Participants were given “homework” to submit entries for worst-case scenarios. They had to be realistic — based on current technologies or those that appear possible — and five to 25 years in the future. The entrants with the “winning” nightmares were chosen to lead the panels, which featured about four experts on each of the two teams to discuss the attack and how to prevent it.

[Read full story here...](#)