

TECHNOLOGY | ARTIFICIAL INTELLIGENCE

The OpenAI Board Member Who Clashed With Sam Altman Shares Her Side

In an interview, AI academic Helen Toner explains her posture in OpenAI's power struggle

By [Meghan Bobrowsky](#) [Follow](#) and [Deepa Seetharaman](#) [Follow](#)

Updated Dec. 7, 2023 4:39 pm ET



Helen Toner is one of the board members who voted to fire Sam Altman from OpenAI. He is back at the company, and she has resigned. PHOTO: JEROD HARRIS/GETTY IMAGES FOR VOX MEDIA

Helen Toner was a relatively unknown 31-year-old academic from Australia—until she became one of the four board members who fired Sam Altman from the artificial-intelligence company he co-founded.

Thrust into the spotlight during the ouster and eventual return of Altman as CEO of OpenAI last month, Toner has emerged as a symbol of tension between AI-safety advocates and those giving priority to technological progress.

Toner maintains that safety wasn't the reason the board wanted to fire Altman. Rather, it was a lack of trust. On that basis, she said, dismissing him was consistent with the OpenAI board's duty to ensure AI systems are built responsibly.

"Our goal in firing Sam was to strengthen OpenAI and make it more able to achieve its mission," she said in an interview with The Wall Street Journal.

Toner held on to that belief when, amid a revolt by employees over Altman's firing, a lawyer for OpenAI said she could be in violation of her fiduciary duties if the board's decision to fire him led the company to fall apart, Toner said.

"He was trying to claim that it would be illegal for us not to resign immediately, because if the company fell apart we would be in breach of our fiduciary duties," she told the Journal. "But OpenAI is a very unusual organization, and the nonprofit mission—to ensure AGI benefits all of humanity—comes first," she said, referring to artificial general intelligence.

Ultimately, Toner and some other board members did resign, clearing the way for Altman's return.

In the interview, Toner declined to provide specific details on why she and the three others voted to fire Altman from OpenAI.

The board said in its initial statement when it fired Altman that he hadn't been "consistently candid" in his discussions with directors. The board didn't provide specifics, frustrating the company's employees and investors, and creating a void in which people have put forth their own theories.

Before his ousting, Altman and Toner had clashed.

In October, Toner, who is director of strategy at a think tank in Washington, D.C., co-wrote a paper on AI safety. The paper said OpenAI's launch of ChatGPT sparked a "sense of urgency inside major tech companies" that led them to fast-track AI products to keep up. It also said Anthropic, an OpenAI competitor, avoided "stoking the flames of AI hype" by waiting to release its chatbot.

After publication, Altman confronted Toner, saying she had harmed OpenAI by criticizing the company so publicly. Then he went behind her back, people familiar with the situation said.

Altman approached other board members, trying to convince each to fire Toner. Later, some board members swapped notes on their individual discussions with Altman. The group concluded that in one discussion with a board member, Altman left a misleading perception that another member thought Toner should leave, the people said.

By this point, several of OpenAI's then-directors already had concerns about Altman's honesty, people familiar with their thinking said. His efforts to unseat Toner, parts of which were previously reported by the New Yorker, added to what those people said was a series of actions that slowly chipped away at their trust in Altman and led to his unexpected firing on the Friday before Thanksgiving.

The board members weren't prepared for the fallout from their decision.

The members, including Toner, were taken aback by staffers' apparent willingness to abandon the company without Altman at the helm and the extent to which the management team sided with the ousted CEO, according to people familiar with the matter.

Toner took her account on social-media platform X private during the height of the crisis.

At one point during the heated negotiations, a lawyer for OpenAI said the board's decision to fire Altman could lead to the company's collapse. "That would actually be consistent with the mission," Toner replied at the time, startling some executives in the room.

In the interview, Toner said that comment was in response to what she took as an "intimidation tactic" by the lawyer. She says she was trying to convey that the continued existence of OpenAI isn't, by definition, necessary for the nonprofit's broader mission of creating artificial general intelligence that benefits humanity at large. A simultaneous concern of researchers is that AGI, an AI system that can do tasks better than most humans, could also cause harm.

"In this case, of course, we all worked very hard to ensure the company could continue succeeding," she added.

OpenAI has an unusual structure where a nonprofit board, on which Toner served, oversees the work of a for-profit arm. The board's mandate is to "humanity," not investors.

In the interview, Toner didn't answer questions about her interactions with Altman. She wouldn't comment on whether she would have done anything differently but said she had good intentions.



Sam Altman was fired as OpenAI's CEO, and then returned. PHOTO: JUSTIN SULLIVAN/GETTY IMAGES

Before he was reinstated, Altman offered to apologize for his behavior toward Toner over her paper, according to people familiar with the matter. Ultimately, he returned to lead the company without following through on that gesture.

Toner is known in the AI-safety world for being a critical thinker who isn't afraid to challenge commonly held beliefs.

Some of Altman's backers, including OpenAI investor Vinod Khosla, publicly expressed derision specifically toward Toner and Tasha McCauley, another former OpenAI board member who voted to fire Altman and is connected to organizations that promote effective altruism.

"Fancy titles like 'Director of Strategy at Georgetown's Center for Security and Emerging Technology' can lead to a false sense of understanding of the complex process of entrepreneurial innovation," Khosla wrote in an essay in tech-news publication the Information, referring to Toner and her current position.

"OpenAI's board members' religion of 'effective altruism' and its misapplication could have set back the world's path to the tremendous benefits of artificial intelligence," he wrote amid the power struggle.

Toner was previously an active member of the effective-altruism community, which is multifaceted but shares a belief in doing good in the world—even if that means simply making a lot of money and giving it to worthy recipients. In recent years, Toner has started distancing herself from the EA movement.

"Like any group, the community has changed quite a lot since 2014, as have I," she said.

Toner graduated from the University of Melbourne, Australia, in 2014 with a degree in chemical engineering and subsequently worked as a research analyst at a series of firms, including Open Philanthropy, a foundation that makes grants based on the effective-altruism philosophy.

In 2019, she spent nine months in Beijing studying its AI ecosystem. When she returned, Toner helped establish a research organization at Georgetown University, called the Center for Security and Emerging Technology, where she continues to work.

She succeeded her former manager from Open Philanthropy, Holden Karnofsky, on the OpenAI board in 2021 after he stepped down. His wife co-founded OpenAI rival Anthropic.

"Helen brings an understanding of the global AI landscape with an emphasis on safety, which is critical for our efforts and mission," Altman said when she joined the board.

The new board members along with returning board member Adam D'Angelo offer a glimpse of the direction OpenAI might be headed. Larry Summers, former Treasury secretary, and Bret Taylor, former Salesforce co-CEO, appear to be more traditionally business-minded than Toner, McCauley and the third board member who was succeeded, Ilya Sutskever, OpenAI's chief scientist.

There are no longer any women on the board, though the company is expected to expand it in coming months.

"I think looking forward is the best path from here," Toner said.

Write to Meghan Bobrowsky at meghan.bobrowsky@wsj.com and Deepa Seetharaman at deepa.seetharaman@wsj.com

Appeared in the December 8, 2023, print edition as 'Ex-Director of OpenAI Explains CEO Clash'.