



BIOGRAPHY

Chief Digital and Artificial Intelligence Office



DR. MATTHEW KUAN JOHNSON

CHIEF OF RESPONSIBLE AI, U.S. DEPARTMENT OF DEFENSE

Dr. Matthew Kuan Johnson serves as Chief of Responsible AI (RAI) for the U.S. Department of Defense (DoD), where his team is based within the Chief Digital and Artificial Intelligence Office (CDAO). As the DoD's lead for Responsible AI, his Division builds the technical tools, assessments, best practices, governance processes, training, and culture within the Department that support the operationalization and implementation of Responsible AI and the DoD AI Ethical Principles – in addition to serving as the Department's primary technical advisor on Responsible AI. Recently, his Division produced and issued the Department's guidance on Generative AI development and use, based on insights from various research initiatives conducted and overseen by his team.

Dr. Johnson is also the Chief Architect of the [DoD's Responsible AI Toolkit and Web App](#), which contains integrated assessments, tools, and a risk management framework for implementing responsible AI and managing risk. His team is currently developing versions of the Toolkit tailored to various use cases, including Generative AI. He serves as the US representative to NATO's Data & AI Review Board (DARB), where he co-chairs the sub-group on Assessments, Toolkits, and Standards; and co-chairs the Risk Management Working Group for the U.S. Government's Chief AI Officer Council, which is tasked with developing the risk management resources needed to comply with the White House's AI policies and associated requirements.

Dr. Johnson's background is in philosophy and cognitive science, having earned a PhD in Philosophy from the University of Cambridge, as well as degrees in Cognitive Science from the University of Cambridge (MPhil) and Yale University (BA). Previously, he consulted for Google AI and was a Research Fellow at the University of Oxford. His areas of specialty are AI-enabled autonomy, generative AI/LLMs, human cognition, human machine teaming, algorithmic bias, and AI Ethics.