

ACEPS PRESENTS



ACEPS
THE AFRICAN CENTRE
FOR EPISTEMOLOGY AND
PHILOSOPHY OF SCIENCE

IMAGINING & BUILDING WISE MACHINES: THE CENTRALITY OF AI METACOGNITION

Prof Igor Grossmann

27 MARCH 2025 @ 11:00-13:00 SAST (UTC+2)

Recent advances in artificial intelligence (AI) have produced systems capable of increasingly sophisticated performance on cognitive tasks. However, AI systems still struggle in critical ways: unpredictable and novel environments (robustness), lack of transparency in their reasoning (explainability), challenges in communication and commitment (cooperation), and risks due to potential harmful actions (safety). We argue that these shortcomings stem from one overarching failure: AI systems lack wisdom. Drawing from cognitive and social sciences, we define wisdom as the ability to navigate intractable problems - those that are ambiguous, radically uncertain, novel, chaotic, or computationally explosive - through effective task-level and metacognitive strategies. While AI research has focused on task-level strategies, metacognition - the ability to reflect on and regulate one's thought processes - is underdeveloped in AI systems. In humans, metacognitive strategies such as recognizing the limits of one's knowledge, considering diverse perspectives, and adapting to context are essential for wise decision-making. We propose that integrating metacognitive capabilities into AI systems is crucial for enhancing their robustness, explainability, cooperation, and safety. By focusing on developing wise AI, we suggest an alternative to aligning AI with specific human values - a task fraught with conceptual and practical difficulties. Instead, wise AI systems can thoughtfully navigate complex situations, account for diverse human values, and avoid harmful actions. We discuss potential approaches to building wise AI, including benchmarking metacognitive abilities and training AI systems to employ wise reasoning. Prioritizing metacognition in AI research will lead to systems that act not only intelligently but also wisely in complex, real-world situations.



Igor Grossmann is a Professor of Psychology at the University of Waterloo, Canada, whose research focuses on the foundations of wise judgment, including how social and cognitive factors, moral reasoning, and emotions shape decision-making and behaviour in changing cultural contexts. He is known for founding and leading major initiatives such as the Social Science Forecasting Collaborative, the International Wisdom Summit, the Futurescape project, and the Wise Judgment Consortium, aiming to redefine decision-making beyond Western-centric models. Grossmann's research has been widely published in top-tier outlets, reflecting his commitment to interdisciplinary work that unpacks the complexities of culture, cognition and social influence. As an elected member of the College of the Royal Society of Canada, he continues to influence the academic community. In addition to his research, Grossmann also co-hosts the "On Wisdom Podcast," bringing scientific insights on social cognition and behaviour to a broad audience.

African Centre for Epistemology
and Philosophy of Science
4th Floor Boardroom @ UJ on Empire
University of Johannesburg

Inquiries: dimphom@uj.ac.za / shened@uj.ac.za

ACEPS: <https://www.uj.ac.za/aceps>

Organisers: Dimpho Moletsane, Shené de Rijk, Veli Mitova

The University of Johannesburg encourages academic debate and discussion that are conducted in a manner that upholds respectful interaction, safety of all involved, and freedom of association as enshrined in the law, the Constitution, and within the boundaries of the University policies. The views expressed during events are expressed in a personal capacity and do not necessarily reflect the views of the University of Johannesburg.

9 INDUSTRY, INNOVATION
AND INFRASTRUCTURE



20 YEARS
— 2005-2025 —

Our Future. Reimagined.