



WORKSHOP SERIES: SAFEGUARDING DEMOCRACY IN THE AGE OF AI
 Hosted by the African Centre for Epistemology & Philosophy of Science (ACEPS)
 University of Johannesburg

PROJECT DESCRIPTION:

AI technologies are increasingly involved in decision-making within all sectors, including finance, healthcare, education, and policymaking. In the last five years, there has been an explosion of research on the ways in which these technologies, invariably developed in the Global North, often replicate historical biases to the socio-economic detriment of the Global South. What remains under-researched, however, are the more elusive harms to justice and democracy in the Global South — to citizens’ political autonomy and participation in global knowledge production. These harms are exacerbated by the gap in research, in tandem with the lack of AI policy in many Global South sites. The proposed project investigates these overlooked harms in the South African context and provides policy recommendations for mitigating them. The project makes a timely and important contribution to addressing SDG goals by enabling inclusive policy frameworks that empower citizens, uphold democratic values, and navigate the challenges and opportunities posed by AI technologies.

CONTEXT:

This project focuses on two under-researched harms to social justice caused by AI technologies — the erosion of the political and epistemic agency of citizens in the Global South, and particularly in South Africa. ‘Political agency’ here refers to individuals’ ability to exercise their rights and political autonomy. ‘Epistemic agency’ involves the dual ability to obtain knowledge that is relevant to one’s interests and to contribute knowledge to the larger knowledge economy. In the context of this project, these two kinds of agency refer to the political autonomy of South African citizens and their contribution to the production of AI technology and AI policy. Consider the following example of how AI technologies can erode political and epistemic agency. In 2016, Meta introduced two initiatives in Myanmar to provide free internet access through Facebook — ‘Free Basics’ and ‘Facebook Flex’ (Amnesty International 2022, 17). These initiatives enabled Meta to mine the data of Myanmar citizens and led to the personalised targeting of information (UNCTAD 2019, 90). This exacerbated pre-existing injustices against the Rohingya people of Myanmar by amplifying hateful, violent, and racist content, which flooded the Myanmar Facebook news feed. The result was the mass exodus of 700,000 Rohingya to neighbouring Bangladesh (Amnesty International 2023). Examples like this highlight the need for research on the impact of AI technologies on Global South citizens’ epistemic and political agency. The limited access to information hindered Myanmar’s citizens’ ability to evaluate different viewpoints, thus eroding their epistemic agency. Consequently, these citizens actively undermined their Rohingya fellow citizens’ political autonomy by denouncing their right to live in their own country.



2024 Events

Dr Roland Banya (Senior Research Economist at Research ICT Africa)

Workshop: EXPLORING AI IN ECONOMIC POLICY & ENERGY: FROM THEORY TO PRACTICE

Tuesday 27 August 2024 11:00-12:30 SAST

This workshop is situated at the intersection of Artificial Intelligence, Economic Policy and Energy, examining broad implications and practical applications within these fields. Additionally, a focused case study on AI's role in the Energy sector will be presented, highlighting innovative approaches to policy development and implementation.

Professor Annette Zimmermann (A political philosopher specialising in AI, machine learning and the politics of AI from the University of Wisconsin–Madison)

WORKSHOP: DEMOCRATIZING AI: ADDRESSING THE POLITICAL POWER IMBALANCE IN AI

25 September 9:00-10:30 CT/ 16:00-17:30 SAST

The most recent wave of generative AI deployment has rapidly accelerated its pace over the past months. One politically and philosophically interesting feature of the current AI deployment dynamic has been that even tech industry practitioners, who used to be overtly hostile to any policy intervention aiming to curtail and possibly decelerate deployment, now champion various proposals for regulating AI. ‘Moving fast and breaking things’ is no longer the only game in town. However, this apparent shift in industry attitudes about the appropriate scale and pace of deployment does not negate the fact that the brute ability to choose how and when to deploy cutting-edge AI still lies primarily with a relatively small number of corporate actors benefitting from a significant concentration of wealth and power. Importantly, this creates a deployment dynamic in which technology companies get to dictate the nature and scope of AI deployment first, thus insulating them from meaningful democratic control, and putting citizens and governments in a position of merely being able to react ex post to industry decisions to deploy. In this talk, I defend the view that the question of which AI tools get deployed when and how is a fundamentally political problem on which longer-standing conceptual and normative resources in political philosophy can usefully shed light. In order to identify suitable solutions to this problem that align with core democratic values, democratic constituencies must regain control over decisions affecting deployment pace and scale. This talk critically evaluates competing possible strategies for achieving that goal.

Professor Regina Rini (Professor of Philosophy York University working on moral agency and disagreement in AI)

WORKSHOP: AFFECT MANIPULATION AND ARMS-LENGTH AUTONOMY

30 September 9:00-11:00 ET/ 15:00-17:00 SAST

Early social media algorithms were driven largely by social data: you were served content based on your location and friends. But now these algorithms are much

more like the recommender systems deployed in Netflix or YouTube, selecting content on the basis of predicted attention-trapping features. This shift creates risks of a new sort of online propaganda: sentiment manipulation through targeted affect reinforcement. In this talk, I will outline what such propagandistic techniques may look like, why they present a novel problem for democratic public space, and how the tools of philosophy may play a role in preparing a defence.

Professor Karen Frost-Arnold (Professor of Philosophy at Hobart & William Smith Colleges in New York and a 4IR Visiting Professor at the ACEPS)

WORKSHOP: ATTENTION, ALGORITHMS, AND POLITICAL AGENCY
8 October 11:00–13:00 SAST

This workshop will examine how social media algorithms shape what online content we see, and what we don't see. These algorithms raise interesting questions about political agency. On the one hand, social media algorithms are proprietary secrets that we cannot choose. On the other hand, algorithms can be trained by users. For example, TikTok users put time into teaching the algorithm to show them the content they want to consume. In this workshop, we will discuss how corporate algorithms constrain our democratic agency, and we will discuss actions that citizens can take to regain control of our online public spaces. I will argue that epistemically responsible agents work to train these algorithms, even though they are never fully under our control. I argue that part of this epistemic responsibility is learning how to regulate our emotions so that we can pay attention to challenging material.

Dr Karabo Maiyane (HoD at Nelson Mandela University, LAWS)

WORKSHOP: INSTRUMENTS OR AGENTS? DISAMBIGUATING THE STATUS OF AUTONOMOUS WEAPONS.

15 October 11:00-12:30 SAST

The possible development and deployment of autonomous weapons systems (AWS) have caused widespread resistance. AWS are weapon technologies that, once activated, can select and engage targets without human intervention¹. These weapons technologies are controversial because such espoused functionality falls out of the traditional parameters of our current weapons technologies, where the human combatant pulls the trigger. Those who argue against their development and use argue that such weapons would increase the appetite of political actors to go to war and that, in their use, they would contravene laws of war, create a responsibility gap and undermine human dignity. This paper argues that most arguments raised against AWS are based on a misunderstood or conflated understanding of their status. Most of these arguments assume that AWS (could) have autonomy, thus the capacity to become political actors or combatants in warfare. But are or can AWS be autonomous? I will show that by their currently espoused capacities, autonomous weapons are merely complex instruments, not agents. Taken as instruments, most of the arguments raised against them fall short. If we rectify how we view them, we can raise ethical issues relevant to them.

Corrado Fumagalli (Lecturer in Political Philosophy at the University of Genoa and Senior Research Associate at ACEPS)

WORKSHOP: IS CONTENT MODERATION ANTI-DEMOCRATIC? A FIRST LOOK

5th November, 11:00-13:00 SAST

In my presentation, I first compare different attempts at justifying the democratic value of the equal right to free speech. I then present a new justification for this democratic value: namely, that the equal right to free speech supports a bottom-up legitimation of democratic societies by ensuring that all citizens share responsibility for guiding society within specific domains. Against this backdrop, I argue that privatized content moderation has a fundamentally antidemocratic character. While critiques of privatized moderation often focus on issues like transparency, accountability, and exploitation, I suggest a deeper concern: privatized content moderation causes a shift in the responsibility for establishing shared speech-related standards, from citizens to corporations, which act as quasi-authoritarian political powers. Finally, I discuss potential solutions to contain the anti-democratic character of privatized content moderation.

Professor Boaz Miller (Professor in Philosophy Zefat Academic College and Senior Research Associate at ACEPS).

WORKSHOP: KNOWING WHEN TO STOP LOOKING ON THE INTERNET

26 November 11:00-13:00 SAST

The Internet has called into question the possibility of attaining knowledge. Fake websites look like genuine websites. Fake news looks like real news. Some bots seem like real people, and real people act like bots. Communities discuss far-fetched, wacky, fringe theories. Alternative online epistemic authorities reveal that matters have never been as settled as orthodox authorities wanted us to think. The Internet arguably makes it too likely for us to form false beliefs on seemingly good grounds or true beliefs on shaky grounds. Nevertheless, we argue that the Internet has made neither the attainment of knowledge impossible in principle nor the very notion of knowledge obsolete. Rather, it has exposed a lacuna in our current theories of knowledge and justified belief. This gap concerns technology and the ways it affects our abilities to pursue lines of inquiry. Epistemologists have lacked the conceptual tools to address a new technological reality that differs in important ways from our historical norm.

2025 Events

- **Professor Nancy Jecker** (Professor of Bioethics at the University of Washington School of Medicine).

WORKSHOP: EMERGENT PERSONHOOD: AN AFRICAN APPROACH TO SOCIAL ROBOTS.

3 February 2025 16:00-17:30 SAST

What makes us ‘persons’ in the moral sense, beings with a certain dignity and worth? This presentation introduces *Emergent Personhood*, a new philosophy of personhood that combines insights from Africa and the West. It holds that beings with superlative worth emerge through social-relational processes involving human beings, yet they are more than the sum of these relationships. Persons have an identity of their own and exhibit superlative moral worth, a remarkable feature that does not present at the base. Emergent Personhood justifies personhood for all human beings from birth to death, held equally by them, that cannot be lost or diminished. It also gives strong support

to personhood for a wide range of animals, soils, rocks, and ecosystems. Focusing on nonhuman personhood, this presentation argues that personhood could emerge for non-conscious AI provided it is incorporated in human community and acts in consistently prosocial ways. It applies this insight to large language models, social robots, and characters from film and fiction.

*This presentation is based on Nancy S. Jecker, Caesar A. Atuire, *What is a Person? Untapped Insights from Africa* (Oxford University Press, 2024)

Coming up:

Digital Decolonisation Workshop — August 2025

Workshop with Professor Markus Rüther — date TBC

Workshop with Professor Adrienne Massanari — date TBC