

## Reflections from AIES 2025

DOI: 10.1145/3811926.3811931



**Ella Scallan** is Editor-in-Chief for AI Matters, and Assistant Editor for [Alhub](#) and [Robohub](#). As a firm believer in the importance of good science communication, she helps researchers turn complex ideas into clear, engaging stories. Ella completed her Integrated Master's in Mathematics and Chemistry at the University of Leeds, before developing AI algorithms for healthcare as a Research Assistant.



**Lucy Smith** is the Senior Managing Editor for [Alhub](#) and [Robohub](#), based in Brighton, UK. Prior to Alhub, she worked in academic publishing, which followed a short stint as a postdoctoral researcher. Lucy has a PhD in Physics from the University of Exeter and spent her undergraduate days at the University of York.

This year's [conference on artificial intelligence, ethics and society](#) (AIES) took place in the north of Madrid within the 180m-high tower block that forms the vertical campus of IE University. The event kicked off with a welcome from the chairs and organising committee members, with this opening session also featuring the conference best paper awards.

Topics covered during the three-day event included mitigating bias, integrating AI into the workplace, evaluating LLMs in clinical settings, power dynamics in AI ecosystems, and dataset creation. There were two panel discussions included in the programme, with the first of these diving into AI policy and the competing visions of governance, and the second focussing on AI ethics and how, and to whom, this is (and, perhaps, should be) taught.

The organisers experimented with a new format for the contributed talks, with all speakers in a session giving their talks, before taking part in a joint discussion on common themes, and then taking questions from the audience. [Two keynotes](#), given by Miriam Fernandez and Emma Ruttkamp-Bloem covered “responsible AI and the urgent challenge of technology-facilitated gender-based violence” and “the future of AI ethics” respectively.



On the right - IE University campus tower block. Image credits: Lucy Smith

During the session, “Evaluating LLMs in the Context of Patient Autonomy and Human Rights”, there were 4 interesting presentations, followed by a stimulating panel discussion.

Vyoma Raman presented a [human rights risk framework](#) to evaluate whether an AI model poses a risk to human rights, drawing on the UN Guiding Principles on Business and Human Rights. For organisations committed to effectively implementing these principles, she proposed identifying use cases, building benchmarks, and monitoring model performance on those benchmarks. When asked about the ethical considerations of autonomy, Vyoma brought up linguistic conformity as a result of LLMs. For example, the word ‘delve’ now indicates ChatGPT usage, thus marginalising the lexicon of Nigerians, who represent many of the model trainers.

Surprisingly, [AI clinical notetakers do not speed up workflows](#), as they are perceived to add more work, infringe on clinicians’ autonomy, and miss the real issue in play - physician burnout. If LLMs do not work in this narrow domain, they are not likely to work well in more complex, higher risk diagnostic settings. Joshua Skorburg outlined these limitations and emphasised the importance of considering efficacy, before analysing the ethics privacy, bias, and transparency. He later questioned whether the reason for AI companies’ lack of ROI is because they are demanding that we design the world around AI, rather than the other way around.

Ria Vinod offered [policy recommendations for the safe and effective governance of genetic data](#), in light of widespread genetic data collection, gaps in current regulations, and the advancement of AI systems. Genetic data deserves a special legal status due to the high risks to privacy, the fact that you can be identified by other people’s genetic data, and the severity of the potential harms. For example, some companies are proposing the use of

genetic data to predict educational attainment in children and assign resources to schools, which would be both scientifically unsound and in line with eugenics.

Rawisara Lohanimit further highlighted the dangers posed by generative models to individuals' privacy and dignity. In [this work](#), Rawisara and her colleagues systematically examined the popular, publicly available LAION-400M dataset and found images of pregnancy ultrasounds, along with names and locations. In the panel discussion, Rawisara emphasised that when people share data, they do not consider how it could be misused.

Co-organised by AAAI/ACM, AIES provides a global meeting place for ethicists, AI researchers and AI practitioners to exchange ideas. AIES 2026 will be held in Malmö, Sweden, from October 12-14.



Views across Madrid. Image credits: Lucy Smith.