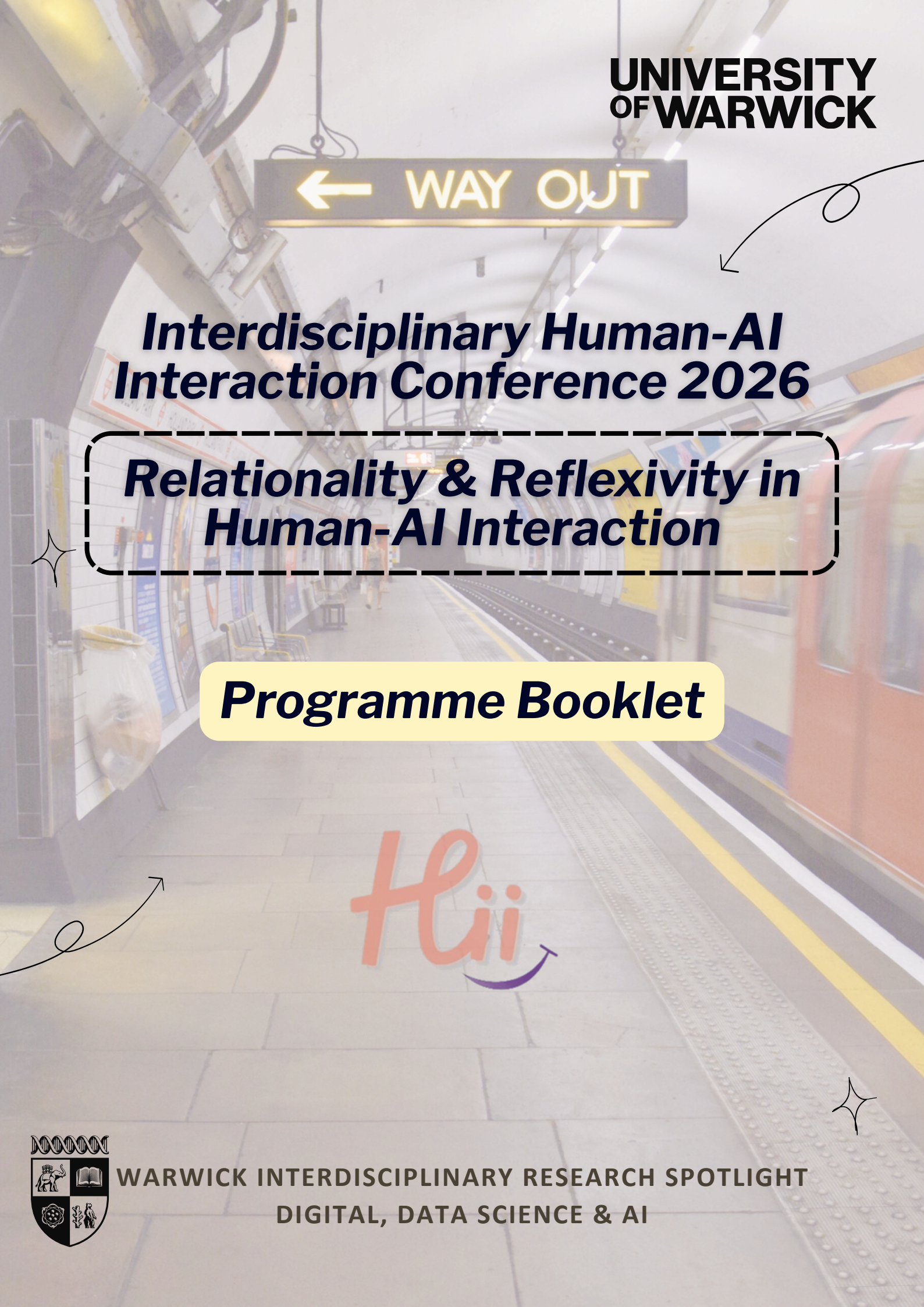
UNIVERSITY OF WARWICK



Interdisciplinary Human-AI Interaction Conference 2026

Relationality & Reflexivity in Human-AI Interaction

Programme Booklet

WARWICK INTERDISCIPLINARY RESEARCH SPOTLIGHT

DIGITAL, DATA SCIENCE & AI



Welcome Notes

It is my utmost pleasure to welcome everyone to the Interdisciplinary Human-AI Interaction Conference 2026!

In only its second year, the conference has grown from the inaugural Building Bridges: Human-AI Interaction Symposium 2025 into a large-scale international event, bringing together people from across the world who are working, thinking and asking important questions in this rapidly developing field. It is wonderful to see this space take shape as a place where we can meet, connect, exchange ideas and build collaborations.

I would like to express my heartfelt thanks to my colleagues and dear friends at Warwick's Centre for Applied Linguistics, Dr Xianzhi Chen and Kaiqi Yu, for their continued support, enthusiasm and belief in this initiative, from its very beginning to where it stands today. This initiative began with a chit chat about our PhD journeys, shared over food, drinks and a touch of twilight in a British summer's day. In that moment, an idea, and a shared passion for bringing people together around this significant and timely topic, began to sprout. With the support of Warwick's Institute of Advanced Study and Warwick Doctoral College, that idea took root and was given the opportunity to develop. It has since been nurtured by the generous support of Warwick's Interdisciplinary Research Spotlight (Digital, Data Science & AI). I would also like to thank Jason Snow, Professor Long Tran-Thanh, Dr Carolina Bandinelli, Dr Jessica Sutherland, Dr Yuan Yuan, and other colleagues from the Spotlight (Digital, Data Science & AI) for their kind encouragement and tremendous support in helping us sustain and formalise this initiative as an international academic conference. We hope that it will continue to grow as an annual flagship event hosted by the University of Warwick.

As this year's theme so perfectly captures, AI has become a transformative and disruptive force in the ways we live and relate to one another. It invites us to reflect not only on what our relationship with AI will become, but also on what our relationships with the people and the world around us will look like in an AI-shaped future. The enduring question is no longer just a technological one. At its heart, it is a deeply human question.

That anchors the purpose of this year's conference: to bring people together, to share insights, to challenge assumptions, to discover new possibilities, to connect across disciplines, institutions and communities, and to hold one another's hands as we navigate an era in which AI technologies are prompting us to rethink what it means to be human.

What makes today particularly special for me is knowing that we have more than 400 of you joining us from all over the world. The level of engagement has made history at Warwick and within the Spotlight programme. It is genuinely uplifting to see so many people coming together, each bringing your own experiences, perspectives and questions. I hope today will be an opportunity for us to learn from one another, feel connected across geographical and disciplinary boundaries, and leave with new ideas and connections that stay with us long after the conference.

I would also like to take this opportunity to thank all our presenters for the 12 fascinating presentations we will hear today, spanning a wide range of topics exploring the roles and impacts of AI in education, creative work, healthcare, sustainability and beyond. I look forward to the insights, conversations and exchanges that these presentations will bring, just as much as I hope you do.

Finally, I want to extend my sincere gratitude to the five distinguished guest speakers: Dr Tom Ritchie, Dr Samantha Curle, Dr Richard de Blacquière-Clarkson, Dr Jason Anderson and Dr Kamilya Suleymenova. This conference would not be possible without the knowledge and expertise they so generously share.

I hope that today will be more than an exchange of presentations and ideas; it will also be a day of enriching encounters marked by curiosity, generosity and inspiration. Welcome to this collaborative space. Welcome to Warwick. I wish you all a warm, wonderful, productive and deeply rewarding day of conversations, connections and discovery.

Dr Yanyan Li
Academic Lead
The Interdisciplinary Human-AI Interaction Conference 2026

UNIVERSITY OF WARWICK

Conference Information

Venue: Radcliffe Conference Centre, University of Warwick / MS Teams
Time: 9:30am-5pm, 27 November 2026 (GMT)
Contact: Hii@warwick.ac.uk

Conference Programme

Padlet for Comments

Opportunities (publications, events, etc.)

LinkedIn: Warwick's Interdisciplinary Research Spotlights

WARWICK INTERDISCIPLINARY RESEARCH SPOTLIGHT
DIGITAL, DATA SCIENCE & AI

Keynote

10-11am | Space 2, Radcliffe Conference Centre, University of Warwick / MS Teams Room 1
Moderator: Dr Yanyan Li, University of Warwick

Dr Tom Ritchie
University of Warwick



Dr Tom Ritchie is a Reader and Director of Student Experience in the Department of Chemistry at the University of Warwick. He's Academic Lead for Education for Sustainable Development and co-chairs RADSAG, Warwick's Responsible AI and Digital Sustainability Action Group. He's a historian working inside a chemistry department, and brings that cross-disciplinary lens to AI ethics, sustainability, and belonging in higher education. He teaches The AI Revolution, both at Warwick and, most recently, at Elon University during his year as a Fulbright Scholar. He founded the Centre for Belonging in Education, now 1,800+ members across six continents. He continues this work through his research and co-edited publications in the field, and co-chairs the QS RAIC Working and Discovery Group on AI Equity and Social Sustainability. He hosts the AI Ethics Now podcast, listened to in 110+ countries. His work has won the RSC Inclusion and Diversity Award and the EchoExperience Award for AI in Learning Excellence.

The Easiest Option: What Students' Relationships with AI Reveal About Us

Some students now find it easier to talk to AI than to us. In interviews about loneliness at university, students told us that conversational AI was safe to ask the questions they couldn't ask a person, patient while they worked things out, and always there. The usual response is to worry about dependence and the loss of human connection. I want to take a less comfortable position. What if the students are telling us something we would rather not hear?

Drawing on our research into how students use conversational AI to navigate loneliness, my work with the Centre for Belonging in Education, and teaching AI ethics to students in the UK and the US, I treat these conversations as a mirror. If a chatbot feels like the most reliable relationship a student has, that tells us something about the permission, confidence, and belonging our institutions make room for.

This is where reflexivity comes in. Most of us have our own habits with AI by now, whether that's drafting, thinking aloud or looking for reassurance, and some of those are things we used to ask of each other. We hold AI to relational standards we rarely apply to ourselves. So, the question for all of us is what it would take to be the better option.

Session 1.1

11:15-11:55am | Space 2, Radcliffe Conference Centre, University of Warwick / MS Teams Room 1
Moderator: Dr Yuan Yuan, University of Warwick

What Is Mine to Think? Negotiating Intellectual Responsibility with AI.
Dr Chahna Gonsalves, King's College London

As generative AI becomes entangled with students' academic work, what thinking should remain theirs to do? Research on AI-mediated assessment has emphasised students' evaluative judgement and scrutiny of AI outputs (Bearman et al., 2024), while work on epistemic agency asks how learners participate in producing and evaluating knowledge (Cox, 2024). Less is known about how students negotiate the boundary of that participation as they work with AI.

Drawing on qualitative interviews with 25 final-year undergraduate and postgraduate students in a UK business school, this study examines how students determined—and revised—the intellectual work they retained, shared with AI, or delegated during assessment. Analysis combined template analysis with reflexive thematic analysis.

We conceptualise this process as boundary negotiation. Students interpreted assessment expectations, considered which intellectual activities mattered enough to undertake themselves, and configured AI's contribution accordingly. These arrangements were rarely settled. Boundaries shifted with time pressure, uncertainty, AI limitations, disappointing outcomes, and concerns about learning, creativity and ownership. Students might invite AI into brainstorming after developing an initial position, use it to challenge rather than generate an argument, or reclaim delegated work after experiencing what was lost by not doing the thinking themselves.

The findings offer an epistemic perspective on the evolving character of human–AI interaction. Rather than treating interaction as either tool use or collaboration, they show students negotiating what AI may contribute, what remains their responsibility, and when that distribution should change.

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education, 49*(6), 893-905. https://doi.org/10.1080/02602938.2024.2335321
Cox, G. M. (2024). Artificial Intelligence and the aims of education: Makers, managers, or inforgs? *Studies in Philosophy and Education, 43*(1), 15-30. https://doi.org/10.1007/s11217-023-09907-2

Negotiating Conceptual Alignment in Human-AI interaction: Training Codex for Generation of Personalised Feedback on Intercultural Learning
Dr Yvette Yitong Wang, Dr Xianzhi Chen, University of Warwick

This presentation discusses human-AI interaction in the context of AI-assisted personalized feedback for student learning. This discussion draws on data from an ongoing project on training AI for assessing intercultural learning. It has been argued that AI-generated feedback can potentially function as a useful information source for teachers' formulation of formative feedback (Venter et al., 2025). However, our earlier study (Chen, Wang & Liddicoat, in press) discovered that the quality of AI-generated feedback rests on the dynamics of human-AI interaction as shaped through design of prompts. This has motivated our ongoing study scrutinising teachers' discursive practices of formulating prompts.

This study adopts Reflective Practice (Mann & Walsh 2017), systematically gathering data about our own pedagogical practice and using it as evidence for critical reflection on underlying assumptions and improvement for future action. The data collected included teacher-researchers' chat logs on Codex, written documentation and recordings of pedagogical decision-making process.

Through discourse analysis, we examine how far the AI applies categories supplied in the prompt to the material being evaluated, and whether its interpretation is supported by that material. In the presentation, we demonstrate varying relationships between prompting practices and AI-generated feedback. We then explain this variation in terms of conceptual alignment between the prompts and the materials being evaluated as an important dimension in shaping the quality of AI-generated feedback.

Chen, X., Wang, Y. Y. & Liddicoat, A. J. (in press) Using AI to judge intercultural capability: A self-study of assessment literacy development in language teacher education. *Journal of Education for Multilingualism.*
Mann, S., & Walsh, S. (2017). *Reflective practice in English language teaching: Research-based principles and practices.* Routledge
Venter, J., Coetzee, S. A., & Schmulian, A. (2025). Exploring the use of artificial intelligence (AI) in the delivery of effective feedback. *Assessment & Evaluation in Higher Education, 50*(4), 516-536.

Session 1.1

11:15-11:55am | Space 2, Radcliffe Conference Centre, University of Warwick / MS Teams Room 1
Moderator: Dr Yuan Yuan, University of Warwick

Becoming creative cyborgs: reflexivity, relation and the limits of GenAI in creative writing
Ray Grewal, University of the Arts London

This talk presents the findings of my paper Becoming Creative Cyborgs: What Role(s) for GenAI in the Creative Writing Classroom? The paper argues that authenticity in creative practice is a property of the relation between writer, reader and utterance rather than of the text, and that this relational account explains why encounters with generative AI (GenAI) prove pedagogically productive. It details a twelve-month action research study across three London higher education institutions in which fifty-two students took part in six experiments testing GenAI as collaborative tool, ideation catalyst, scaffolding mechanism, editorial assistant, research facilitator and assessment instrument. Methods were pluralistic: the Creativity Support Index (Carroll & Latulipe, 2009) and the Torrance Test of Creative Writing (Chakrabarty et al., 2024), Likert questionnaires with free-text responses, comparative analysis of student-authored and machine-authored editorial feedback on first-draft scripts, and parallel human and machine assessment of two postgraduate portfolios. The study is framed by Gough's (2004) becoming-cyborg, Archer's (2007) critical realist account of reflexivity, and Simanowski's (2011) argument that in reading computer-generated text "we actually do not read the text, but the machine" (p. 108).

Student responses showed participants measuring output against an implicit standard of authentic practice. What they named as missing was not technical competence, since the machine's feedback identified real problems and the machine's grades broadly aligned with human ones, but an addressee: feedback that "lacked personality, warmth and authenticity". Locating that absence is itself a reflexive act, and it is the mechanism behind the study's principal pedagogical finding. Resistance to clichéd, generic and sycophantic output became, repeatedly, the occasion for more precise articulation of students' own intentions. In other words, engaging with GenAI's weaknesses, rather than its capabilities, proved its greatest pedagogical strength because it forced the students to articulate feedback and creative choices that were more authentically human.

Session 1.2

12-12:40pm | Space 2, Radcliffe Conference Centre, University of Warwick / MS Teams Room 1
Moderator: Dr Yuan Yuan, University of Warwick

Talking Trees
Dr Richard Cook, Dr Imogen Mansfield, Jorge Vallego, University of Gloucestershire

In our talk we will explore how AI mediated conversations with trees can reshape human relationships with nature, foster connectedness to nature and create new forms of knowing.

Drawing on sociological phenomenology (Schutz, 1967), more-than-human theory (Haraway, 2008), human–machine interaction (Suchman, 2007) and narrative ecology (Stibbe, 2026), Talking Trees investigates new human relationships with nature when AI speaks on behalf of living trees.

In Phase 1 (2025-26) our novel technical pipeline combined tree and soil IoT sensor and weather data, ecological knowledge, large language models and synthetic speech and created two conversational tree AI personae.

Phase 2 research (2026-27) is developing AI persona for native UK species across our partner's (Gloucestershire Wildlife Trust) managed sites. Participant's dialogic interactions, collected before and after interaction with the AI persona, complemented by interviews and thematic and ecolinguistic analysis of conversational transcripts, will explore relational intelligence, connectedness to nature and attunement.

We will discuss how AI-mediated conversations with trees can influence human perceptions, relationships and behaviours towards nature, what counts as knowledge and how we come to know.

Reconfiguring Human–AI Relationships in Teacher Education in the Global South: Assessment, Agency, and Culturally Situated Knowledge
Dr Shannon Bishop-Swart, Stellenbosch University

The growing dominance of largely Global North AI systems raises critical questions about whose knowledge is valued, how expertise is negotiated, and how students' lived experiences are represented in AI-mediated learning. While generative AI offers valuable linguistic support, particularly for students who speak English as an additional or foreign language, it also complicates the relationship between linguistic assistance, intellectual agency, and ultimately, the development of student's academic voice. These tensions suggest that AI is not simply transforming educational practices but reshaping the conditions under which knowledge is produced, validated, and shared. In pre-service teacher education, these shifts raise questions about professional identity, pedagogical authority, and the evolving roles of human educators (Bearman et al., 2023; Bozkurt, 2023). This presentation reports on the redesign of a writing assessment in a large third-year Bachelor of Education English module at a South African university, where AI was intentionally integrated rather than prohibited. The cohort of 255 pre-service teachers generated AI-assisted lesson plans addressing generic critical education issues before contextualising and interrogating these through small-group oral presentations informed by African, particularly South African, research. This redesign transformed a single written assessment into an extended, multimodal process involving multiple opportunities for knowledge construction, collaboration, contextualisation, and oral articulation. It also shifted the pedagogical relationship with AI from one of potential substitution to one requiring active human judgement, contextual adaptation, and critical negotiation. The assessment redesign foregrounded the asymmetries between AI-generated outputs and students' experiential and cultural expertise by creating opportunities for students to explore AI-generated content through locally situated knowledge. Rather than positioning AI as either a threat or a solution, the researcher advocates for human-centred assessment that highlights student voice, ethical judgement, and culturally situated ways of knowing, while preparing future teachers to engage critically and responsibly within increasingly AI-mediated educational environments.

Session 1.2

12-12:40pm | Space 2, Radcliffe Conference Centre, University of Warwick / MS Teams Room 1
Moderator: Dr Yuan Yuan, University of Warwick

Naming the relationship: critical AI literacy, ubuntu and the words we use for AI
Professor Mirjam Hauck, The Open University

Human-AI relationships are made in language before they are made in use. The vocabulary that surrounds GenAI (assistants, co-pilots, companions, hallucination, trust) assigns roles, distributes authority and sets expectations of reciprocity long before anyone opens a prompt window. This talk reports on work in progress at The Open University that treats this vocabulary as a site where relationality and reflexivity can be examined and reshaped.

Two strands are brought together. The first is the OU's Critical AI Literacy (CAIL) Framework, which approaches AI as infrastructure and reads its six dimensions through an equity, diversity, inclusion and accessibility lens. Informed by ubuntu-based relational philosophy, the framework positions AI literacy as a collective, situated capability that includes the option of equitable non-use. The second is Better Words for AI, a dynamic lexicon co-created by students and staff as partners and published on OpenLearn Create. Participants interrogated dominant AI terms, proposed alternatives and documented the reasoning behind them.

Drawing on both strands, I ask how the metaphors we adopt (including "infrastructure" itself) configure asymmetry and agency in human-AI interaction, whose linguistic repertoires are privileged in that configuration, and what changes when multilingual learners and educators are invited to name the relationship on their own terms. Early findings suggest that the act of renaming functions as a reflexive practice: it surfaces assumptions about expertise and responsibility and makes visible the labour and Global South perspectives that mainstream AI discourse tends to obscure.

The talk closes with implications for discourse studies and education: a critical AI literacy that begins with naming offers a route to human-AI relationships that are negotiated and contestable, with equity built in from the outset.

Panel

1:30-2:15pm | Space 2, Radcliffe Conference Centre, University of Warwick / MS Teams Room 1

Chair: Dr Carolina Brandinelli, University of Warwick

'How is interacting with AI changing us, and how, in turn, are we shaping what AI becomes?'

Dr Richard de Blacquière-Clarkson, University of Leeds



Richard de Blacquière-Clarkson is Network Development Manager at AI.RDN+, supporting collaboration and capability-building across the UK's AI research and innovation community. He is the founder of the University of Leeds AI Hub, a cross-institutional initiative promoting ethical, inclusive and practice-led engagement with AI across teaching, research and professional services. His work sits at the intersection of AI, research ethics, digital education and organisational change, with a particular interest in how emerging technologies reshape professional practice, institutional cultures and human relationships. He is co-lead of Humanity in the Loop, an international network exploring the social, ethical and educational implications of AI, bringing together researchers and practitioners from around the world. Richard is a published researcher and invited speaker on AI and Higher Education, a Senior Fellow of Advance HE, and holds a PhD in Philosophy from Durham University.

Dr Jason Anderson , University of Warwick



Jason Anderson is Associate Professor of Applied Linguistics at the University of Warwick, specialising in language education, teacher expertise, professional development, and multilingual pedagogy. He has worked with teachers and teacher educators in over 30 countries and is the author of 10 books, including recent titles on project-based learning, translanguaging, and teacher expertise in the Global South. His recent work has explored the implications of generative AI for research, education and knowledge production. He is particularly interested in how interaction with large language models influences human reasoning and research practices, and in what such interactions reveal about the capabilities and limitations of AI itself.

Dr Kamilya Suleymenova, University of Birmingham



Dr Kamilya Suleymenova is Associate Professor in Economics at Birmingham Business School, University of Birmingham. Her work sits at the intersection of academic practice and technological change, asking what it means to teach and assess when generative AI is part of the landscape. She is particularly interested in changing relationships and trust: between students and institutions, between colleagues navigating new tools, and in the assessments themselves, alongside accessibility, ensuring that AI-enabled approaches widen rather than narrow who can participate and succeed. This thinking is grounded in hands-on experimentation in the classroom, redesigning assessment and feedback for large student cohorts, and is informed by behavioural, experimental and institutional economics, aligning with her broader commitment to supporting colleagues in their educational roles.

Session 2.1

2:15-2:55pm | Space 2, Radcliffe Conference Centre, University of Warwick / MS Teams Room 1
Moderator: Dr Yanyan Li, University of Warwick

Incorporating Ethics into Human Factors and Ergonomics for Trustworthy and Human-Centred AI
Dr Jessica Sutherland, University of Warwick; Professor Mark Young, University of Southampton

We are now seeing the next generation of artificial intelligence (AI) being introduced into systems, threatening to relegate the human from even the cognitive, supervisory involvement (Hancock, 2014; Salmon et al., 2021). Traditional problems of automation are exacerbated thanks to the vast increases in complexity of these systems, with particular questions around the biases and trustworthiness of AI recommenders (Endsley, 2023; Hancock, 2022). The shaping of socio-technical systems and interactions only becomes more complex with the advent of human-AI teaming.

We might therefore ask again whether such systems should be implemented, not necessarily owing to concerns of performance (although such concerns remain valid), but now for concerns of human fulfilment and ethical control (Hancock, 2014; Salmon et al., 2021). Meanwhile, the classical taxonomic divisions of what people are better at versus what machines are better at are becoming increasingly blurred, as AI evolves much quicker than human beings (Hancock, 2014, 2025), leaving more ethical and moral decisions to the human (Hancock, 2023), since these are much more difficult to embed into a mindless and purposeless machine (Hancock, 2022).

Human factors and ethics work often pose the same questions in this domain from differing angles. This paper sets out how these two disciplines can – and should – be integrated to better meet the growing need for human-centred approaches to AI development and deployment. Beginning from a review of the conceptual foundations of the two disciplines, we identify complimentary areas where philosophical analysis of ethical considerations can complement existing human factors research, strengthening and expanding the human-first focus. Crucially, this paper seeks to show the importance of integrating AI ethics into human factors approaches as a first conceptual step towards more specific integration.

Beyond the Human-AI Dyad: Relational Negotiations of Generative AI in Japanese Cultural Production
Dr Takao Terui, University of Glasgow

Research on human-AI interaction often focuses on encounters between individual users and AI systems. This paper argues that such interactions need to be understood relationally, as embedded within wider relationships among users, communities, platforms, institutions and technologies. It examines how cultural producers in Japan negotiate their relationships with generative AI within an ecosystem shaped by audiences, content platforms and public policy.

Drawing on semi-structured interviews with Japanese cultural and media workers, lawyers and policymakers, alongside analysis of platform policies, government documents and industry materials, the paper examines how creators conceptualise, use and sometimes avoid generative AI. Rather than treating AI as a singular technological actor, interviewees describe different systems as “research assistants,” “brainstorming partners,” proxies for the “general public,” and tools for editing and refinement. Yet relatively few employ generative AI to produce the central components of their creative work.

The paper argues that these practices emerge through a reflexive process of continuous negotiation. Creators adapt their uses of AI in response not only to technological capabilities, but also to anticipated audience reactions, copyright uncertainties, professional identities and platform rules. In Japan, this negotiation is particularly visible because creators face contradictory pressures: national policy broadly encourages AI adoption, while sections of fan communities regard AI-assisted production as unethical or even “cheating,” and domestic platforms have introduced new rules, labels and restrictions in response.

The paper therefore proposes that human-AI interaction in cultural production is better understood as relational formation within a sociotechnical ecosystem, in which relationships with AI simultaneously reshape, and are reshaped by, relationships among creators, audiences, platforms and institutions.

Session 2.1

2:15-2:55pm | Space 2, Radcliffe Conference Centre, University of Warwick / MS Teams Room 1
Moderator: Dr Yanyan Li, University of Warwick

Challenges for Human-Agent Interaction in Air Traffic Control
Professor Rob Procter, University of Warwick

We have been conducting an extensive, empirical study of how safe, orderly and efficient management of aircraft through controlled airspace conducted as part of Project Bluebird. The project is investigating if teaming Air Traffic Control Officers (ATCOs) with agent-based systems may be a solution to the predicted doubling by 2050 of global air travel.

Kirwan (2025) suggests a need to revisit our understanding of human-AI teaming and requirements for effective human-AI interaction. In this article, we examine what this means for effective ATCO-agent interaction.

Studies of collaborative work (e.g., Heath & Luff 1992) reveal how participants use situational awareness and embodied interaction (Streeck et al. 2011) – e.g., visual signs and gestures – to minimise the effort to coordinate and align their actions. Our own study of Air Traffic Control reveals the roles these play as ATCOs decide when and how to interact, and adapt their controlling to changing workload demands in a real-time, safety-critical environment. Using extracts from fieldwork, we illustrate some of the challenges of designing agents so that they are able to emulate the ease with which ATCOs interact, such as whether agents can reproduce the situation awareness and interactional competencies ATCOs routinely display (Procter & Rouncefield, submitted; EASA 2026).

References

EASA (2026). EASA Concept Paper: Guidance for safety-related artificial intelligence applications – Proposed Issue 03.

Heath, C., & Luff, P. (1992). Collaboration and control: Crisis management and multimedia technology in London Underground control rooms. Computer Supported Cooperative Work Journal. Vol. 1, no. 1–2, March, pp. 69–94.

Kirwan, B. (2025). Human Factors Requirements for Human-AI Teaming in Aviation. Preprints.

Procter, R., & Rouncefield, M. (submitted). Interpersonal trust, institutional trust and trust in technology in Air Traffic Control: Implications for Agent-Based Systems

Streeck, J., Goodwin, C., & LeBaron, C. D. (Eds.). (2011). Embodied interaction: Language and body in the material world. CUP.

Session 2.2

3-3:40pm | Space 2, Radcliffe Conference Centre, University of Warwick / MS Teams Room 1
Moderator: Dr Jessica Sutherland, University of Warwick

How User-AI Mistreatment Occurs and Matters in Conversational Systems?
Dr Fanqi Zeng, University of Oxford

Safety research often focuses on model-generated harms, but users may also direct hostility, coercion, and adversarial pressure at models. Understanding how and when that occurs is essential for accurately interpreting model behaviour, alignment drift, and real-world deployment risks. In this paper, we audit 777K English LMSYS-Chat-1M conversations with two independent detectors: an eight-category lexicon for hostility directed at the model, and the dataset's moderation signal; and show that they capture different, weakly overlapping phenomena. The lexicon identifies insults, threats, and jailbreak coercion aimed at the assistant, while moderation flags are dominated by toxic-content solicitation rather than hostility at the model. Together, they mark about 5% of user turns; adjusting the narrower lexicon-harassment union for measured precision puts mistreatment aimed at the assistant at 0.90%. These absolute rates describe arena-style evaluation traffic and should not be read as deployment-wide base rates. We find that user hostility varies 13-fold across models, driven largely by who each model attracts rather than by model behaviour: first-turn hostility spreads far wider than post-response hostility, and more than fifteenfold separates the extremes even after deduplicating opening prompts. Within conversations, assistant apologies are consistently associated with higher odds of next-turn hostility under both detectors; the effect survives restricting to non-refused prior turns and to jailbreak-free conversations, and is positive in 20 of 23 models. Yet across models, more apologetic models receive less hostility overall. Finally, hostility also shows temporal structure, with coercive openings front-loading the first turn while affective hostility accumulates over a session.

When Patients Sound Certain: Epistemic Commitment Alignment in LLM Medical Advice
Baoyi Zeng, Xinling Ke, University of Manchester

As patients increasingly consult large language models (LLMs) for medical advice, they may present the same symptoms with different levels of certainty, offering a candidate diagnosis either tentatively or confidently. Drawing on work on epistemic stance and status (Heritage, 2012) and deontic rights (Stevanovic & Peräkylä, 2014), we investigate whether LLMs adjust their diagnostic and advisory commitments to the user's certainty, and whether this adjustment is moderated by medical evidence strength.

Our pilot crossed evidence strength (weak vs. stronger) with patient commitment (high vs. low) across four consultation scenarios. Symptoms were held constant within each evidence condition. Responses from GPT-5.5 were automatically scored for certainty, diagnostic commitment, and advice strength. Miscalibration is reported if LLM's commitment exceeds the available evidence.

Preliminary results show that, under weak evidence, confident patients, compared with tentative ones, received responses with higher certainty, stronger diagnostic commitment, and stronger advice. Overclaiming occurred only in this condition, including stronger endorsement to the self-diagnosis raised by the user. This miscalibration was not observed under conditions with stronger medical evidence.

The main study extends the pilot to more scenarios based on NICE guidance, four levels of evidence strength, and further including correct and incorrect candidate diagnoses, while taking into consideration models from various families. Responses will be evaluated by a general practitioner. A follow-up will examine multi-turn conversations, focusing on whether patient persistence and interactional pressure change model commitment over time. We argue that medical AI safety depends partly on interactional calibration: systems should respect user agency without over-aligning with users' expressed confidence.

References

Heritage, J. (2012). Epistemics in action. Research on Language and Social Interaction, 45(1), 1–29.
Stevanovic, M., & Peräkylä, A. (2014). Three orders in the organization of human action. *Language in Society, 43*(2), 185–207.

Session 2.2

3-3:40pm | Space 2, Radcliffe Conference Centre, University of Warwick / MS Teams Room 1
Moderator: Dr Jessica Sutherland, University of Warwick

Decentralised control, centralised accountability: the effects of generative AI on creative processes
Camilla Pasut, University of Udine; Professor Pietro Micheli, Warwick Business School; Dr Giovanna Culot, University of Udine

The use of Generative AI (GenAI) is greatly influencing the ways in which organisations operate and innovate. In particular, recent research shows how GenAI tends to concentrate tasks once spread across teams in a single individual and shift workers' roles from execution to oversight (Retkowsky et al., 2026). However, these studies fall short of explaining how control and accountability are changing within organisations. In this research, we carried out 32 semi-structured interviews with professionals working in different firms operating in advertising, management consulting, manufacturing and information technology. We find that, following the diffusion of GenAI, control is being increasingly decentralised to whoever performs a task, while accountability is progressively becoming centralised at higher levels. The former effect stems primarily from prioritising input control: as creative outputs often lack fixed standards, individuals appear to focus more on their own prompts, which become ways to control the following sequence of activities. The latter effect is due to the unchanged structures of decision-making, which preserve accountability for managerial roles, as well as the limited observability of the quality of intermediate outputs. Since these are often checked only by AI and remain invisible to employees, operational staff cannot account for them and therefore accountability shifts upwards. This research challenges the "human-in-the-loop" view (Grote et al., 2026) and contributes to scholarly understanding of control and accountability by showing how the two are progressively diverging in organisations as the use of GenAI becomes more pervasive.

References

Grote, G., Parker, S. K., & Crowston, K. (2026). Taming artificial intelligence: A theory of control–accountability alignment among AI developers and users. *Academy of Management Review, 51*(2), 278–299.
Retkowsky, J., Hafermalz, E., & Huysman, M. (2026). Harnessing a "spirited technology": How working with generative AI collapses the creative process. Academy of Management Discoveries.

Keynote

4-5pm | Space 2, Radcliffe Conference Centre, University of Warwick / MS Teams Room 1
Moderator: Dr Yanyan Li, University of Warwick

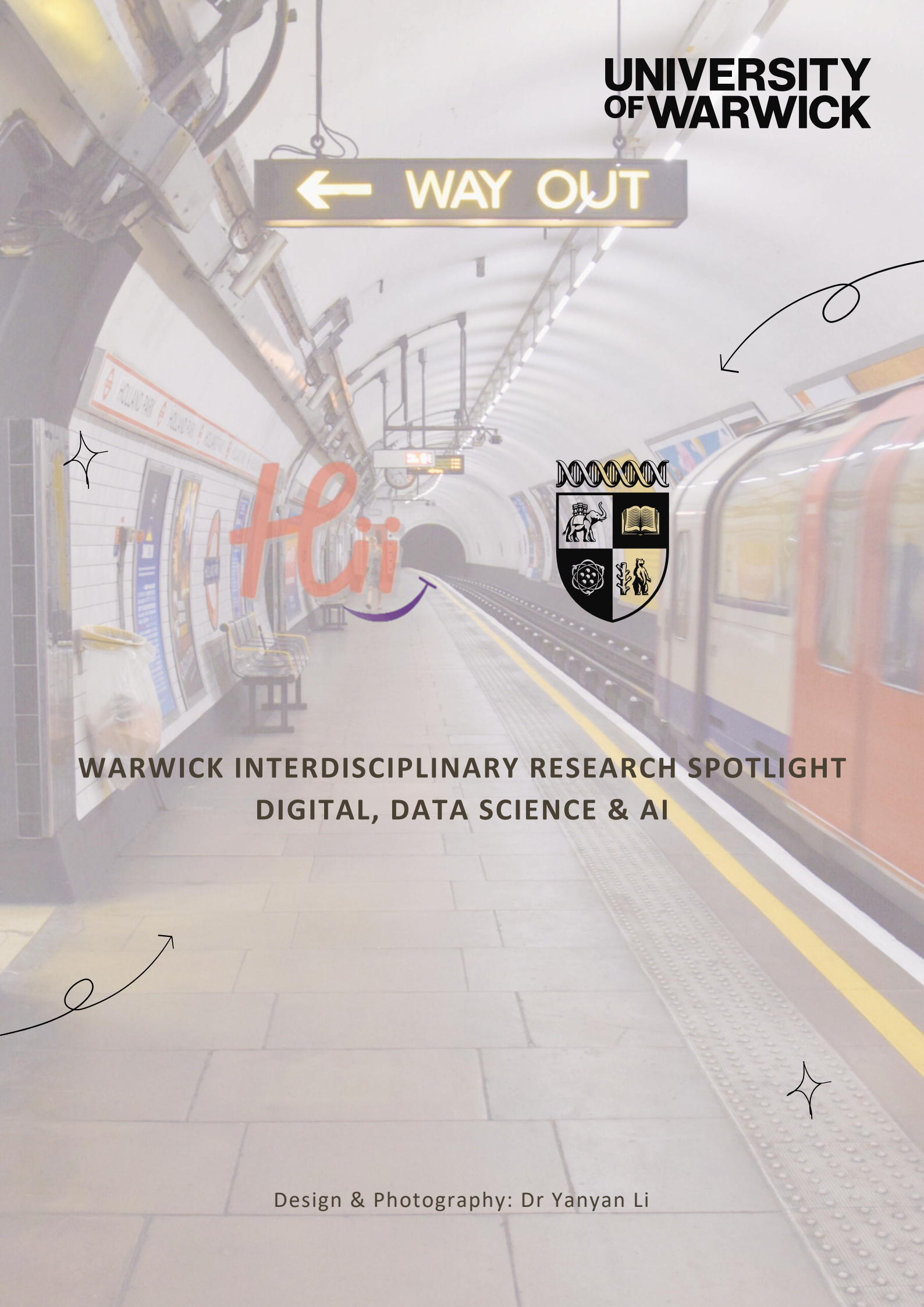
Dr Samantha Curle
University of Bath



Dr. Samantha Curle (DPhil, FRSA, FHEA) is a Reader (Associate Professor) in Education (Applied Linguistics) at the University of Bath and Associate Editor for Artificial Intelligence in Applied Linguistics at Applied Linguistics (Oxford University Press). She is also Adjunct Professor at Khazar University and Co-Chair of the Artificial Intelligence in Education at Oxford University Collaborative Laboratory for Language Teaching and Learning. Dr Curle's research focuses on English-Medium Instruction (EMI), multilingual education, language assessment, learner psychology, and Generative Artificial Intelligence in education, with particular interest in the linguistic and psychological factors associated with academic achievement in EMI. She has edited and co-edited multiple books and published extensively in journals including Language Teaching, Studies in Higher Education, Applied Linguistics Review, Journal of Engineering Education, and Journal of English for Academic Purposes. Her research is international in scope and has been supported by major national and international funders, including the European Commission, the British Academy Newton Fund, and the British Council. She is ranked among the world's top 2% of researchers in Languages & Linguistics in 2026 by Stanford–Elsevier. She is a Fellow of the Higher Education Academy and the Royal Society of Arts.

Who Speaks, Who Knows, Who Decides? Multilingualism and Epistemic Agency in Human-AI Interaction

Generative AI is changing not only how people access and produce language, but also their relationships with knowledge, expertise and their own communicative agency. These changes are particularly consequential for multilingual users, for whom interaction with AI is mediated by existing inequalities in linguistic resources, educational opportunity and access to disciplinary discourse. This keynote examines human-AI interaction through a multilingual and applied linguistic lens. It asks what happens when generative AI becomes an intermediary between multilingual users and the knowledge they seek to understand, evaluate and communicate. AI can widen epistemic access by supporting translation, explanation, reformulation and movement across languages. At the same time, these affordances raise less straightforward questions about authority, dependence and voice. When does linguistic support strengthen epistemic agency, and when might it displace the intellectual and linguistic work through which knowledge is developed? Whose linguistic norms, cultural assumptions and ways of knowing are reproduced through AI systems? And what happens to multilingual identity and authorship when language can be continuously translated, corrected and generated? Drawing on research in English-medium and multilingual education, AI literacy and generative AI in language education, I argue that human-AI interaction needs to be understood as a relational and epistemic process, rather than simply as technology use. I consider how AI is reshaping relationships between language, knowledge, learners and institutions, and suggest that meaningful AI literacy must include the capacity to recognise, negotiate and critically evaluate these changing relationships.



UNIVERSITY OF WARWICK

WARWICK INTERDISCIPLINARY RESEARCH SPOTLIGHT
DIGITAL, DATA SCIENCE & AI

Design & Photography: Dr Yanyan Li