

My general research area is Natural Language Processing (NLP), a branch of artificial intelligence (AI) that enables computers to understand, interpret, and generate human language to communicate with humans. My research aims to build human-grounded NLP systems that represent diverse human meanings and adapt reliably to real-world needs, drawing on a socially and cognitively informed approach that spans NLP, cognitive science, software engineering, and healthcare research.

My research is driven by understanding how meaning varies across people, developing computational representations that capture this diversity, examining how language technologies learn, reproduce, or amplify existing biases, and translating these insights into systems that better serve people in practice. Across these themes, my goal is to make NLP systems **cognitively grounded and socially aligned, faithful in its representations, and adaptive in its use**. These themes form a connected research ecosystem, as shown in Figure 1 and elaborated below.

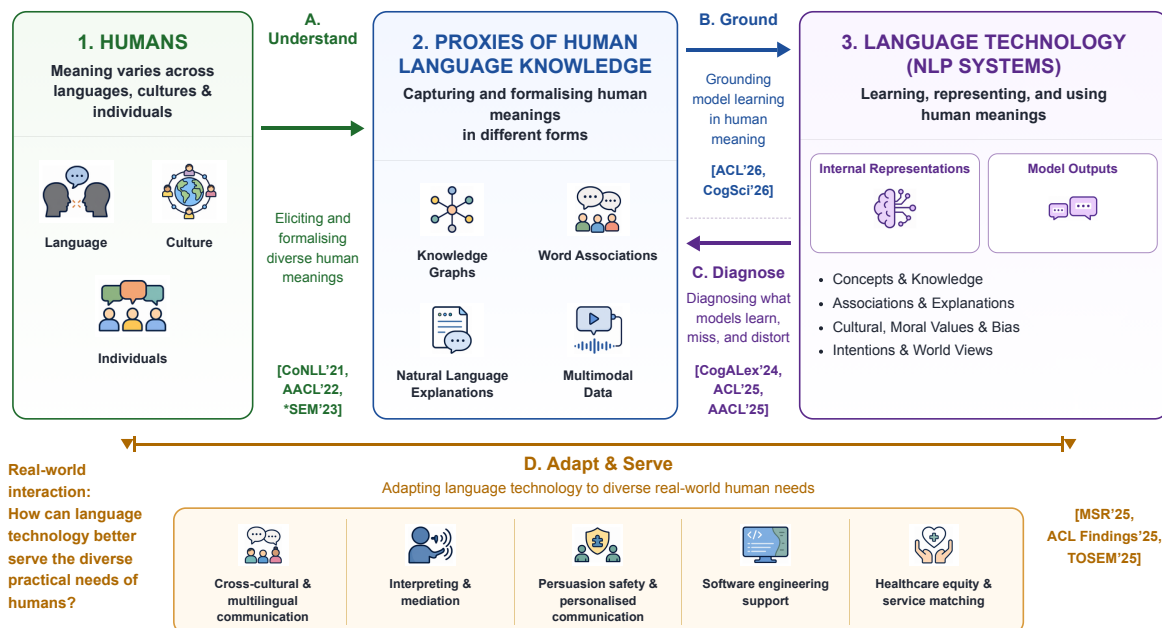


Figure 1: My research connects human, data, technology, and real-world perspectives.

Theme 1: Meaning Variations in Concepts. *How do people understand and interpret meaning within and across languages and cultures?* For example, in the task of free word associations, given the word “smile”, what is the first word that comes to your mind? Is it *face, happy, camera, cry*, or something else? People might give different answers. My research examines how meanings, values, and communicative intentions vary across individuals, languages, and cultures, and what shapes this variation. I curate datasets from humans to capture such variation and develop computational methods to analyse it, aiming to understand the cognitive and social processes underlying language use. This builds on my prior work eliciting natural-language explanations to uncover the reasoning behind people’s free associations [AACL'22; CogSci'23; CogALex'24; CogSci2026] and extends to my ongoing work disentangling conceptual meaning from thematic, contextual, and culturally situated associations across languages.

Theme 2: Cognitively-informed NLP. *How can implicit commonsense knowledge be captured to improve NLP models?* Building on insights from cognitive psychology, my PhD research formalised human semantic memory as a computational resource through word associations, natural-language explanations, and structured knowledge representations. This work showed how cognitively grounded resources can complement existing knowledge graphs extracted from text-corpora

and improve common-sense reasoning in NLP systems [CoNLL’21; AACL’22; *SEM’23]. This body of work formed the core of my PhD thesis, which was awarded **Pass without Amendment** by both examiners. As one examiner praised: “*Given the importance of modeling commonsense knowledge in natural language processing and in artificial intelligence in general, the thesis makes unique contribution to the field. Overall, it is a thesis of very high quality. I recommend Pass without any reservation.*” It established the foundation for my continuing research on multilingual and cross-cultural meaning representations and alignment.

Theme 3: Cultural and Value Alignment. *To what extent do language technologies capture human meanings, and what do they learn, miss, or distort?* Since completing my PhD, I have established an independent research direction on cultural and value alignment in language models. I examine both *internal representations* and *model outputs* to identify where models diverge from human semantic structures and cultural values, and develop methods to reduce these gaps. I have extended my earlier work on commonsense knowledge into cultural and moral alignment [ACL’25; ACL’26 (CORE A*)]. This line of research has attracted both internal funding from the University of Melbourne through a faculty collaboration (**AU\$30K**, 2025) and external funding from Google Australia (**AU\$80K**, 2026).

Theme 4: NLP-driven Practical Applications. *How can language technology better serve diverse human needs in real-world?* I translate NLP methods into adjacent disciplines where they can deliver direct practical impact, working closely with domain researchers and practitioners. One direction is **NLP4SE** (NLP for software engineering), a continuing collaboration with the SE research group in UoM since 2024. I apply NLP to develop trustworthy automated code review systems to improve software reliability (which is becoming increasingly necessary as large language models accelerate code generation). This work has produced five publications across NLP and SE, including ACL’25, AACL’25, MSR’25, SCAM, and **TOSEM**, and has been featured in keynotes at SE conferences [ASE’25 and MSR’26]. Another direction of my research is **NLP4Health**: how can we address structural inequalities in access to mental health services for people from diverse linguistic, cultural, and geographic backgrounds? I currently lead the development of a service database and matching tool within the ALIVE National Centre’s *Whose Care Is Left Behind?*, an AU\$4M MRFF-funded project. Together, these projects demonstrate how I translate NLP expertise into domain partnerships, research outputs, and technologies addressing concrete human and societal needs.

Research Agenda. Over the next three to five years, I will continue to deepen my research program on **cultural interpretation in AI systems**, connecting my work on human meaning, model alignment, and real-world multilingual communication. As AI becomes embedded in everyday communication, it may bridge surface language barriers, but not necessarily cultural ones. We still know little about how cultural biases in AI are represented and directly shape real-world interactions or reinforce inequalities. Addressing this gap is therefore both urgent and consequential.

Addressing this challenge requires an integrated approach across three interconnected layers. This includes **(1) the human layer**, examining how people interpret meaning across languages, cultures, and social contexts; **(2) the model layer**, studying how language models internally represent and interpret cultural meanings and values; and **(3) the mediation layer**, developing AI interpreting systems that bridge linguistic, cultural, and pragmatic gaps in healthcare, public services, and other high-stakes contexts. My ongoing research focuses on the first two layers. Over the next three to five years, I will prioritise the mediation layer to ground my research in real-world settings where communication failures have direct consequences for access, understanding, and equitable outcomes.