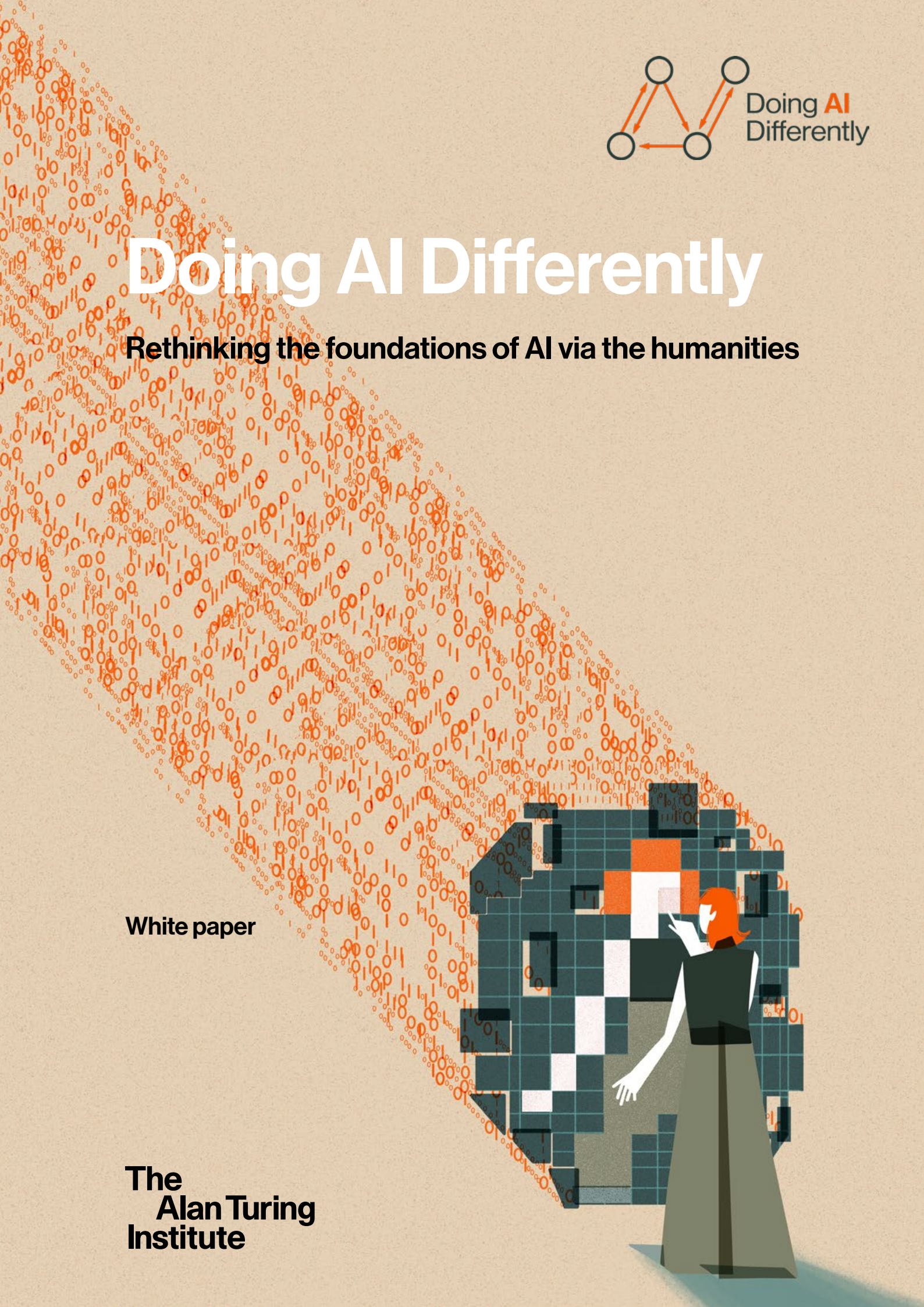
Doing AI Differently

Doing AI Differently

Rethinking the foundations of AI via the humanities



White paper

The Alan Turing Institute

Contents

Contents 2

Authors 3

Executive summary 5

Preface 9

1. Why? 10

1.1 Introduction 10

1.2 Transformative shifts in AI: challenges and opportunities 11

1.2.1 The qualitative turn 11

1.2.2 The homogenisation problem 12

1.2.3 The transformation of human cognition 13

1.3 Articulating the problem space: a transdisciplinary endeavour 14

2. What? 16

2.1 Objectives 17

2.2 Interpretive technologies 18

2.3 Alternative architectures for AI 21

2.4 Human-AI ensembles 24

2.5 Building an interdisciplinary community 27

2.6 Strategic positioning 28

2.7 Real world case studies 29

2.7.1 Sustainability case study 29

2.7.2 Healthcare case study 30

2.7.3 Engineering design case study 31

2.8 Navigating risks and unintended consequences 32

3. How? 33

3.1 Methodology and community engagement 34

3.2 Workstreams 35

3.2.1 Workstream logic table 36

3.2.2 Detailed workstream descriptions 37

W1: Develop interpretive AI foundations 37

W2: Expand AI design pathways through interdisciplinary insight 37

W3: Enable human-AI ensemble systems 38

W4: Build talent, capacity, and cross-sector pathways 38

W5: Establish global knowledge infrastructure 39

3.3 Implementation mechanisms 40

3.4 Barriers and risks 42

3.5 Funding model 43

3.6 Success metrics and timeline 44

3.7 Why we believe this can succeed 46

4 Collaborative next steps 47

Signatories 48

Authors

Lead authors:

Drew Hemment (The Alan Turing Institute & University of Edinburgh)
Cody Kommers (The Alan Turing Institute)

Contributing authors:

Ruth Ahnert (Queen Mary University London)
Maria Antoniak (University of Colorado)
Glauco Arbix (Universidade de São Paulo)
Vaishak Belle (University of Edinburgh)
Steve Benford (University of Nottingham)
Alexandra Brintrup (The Alan Turing Institute & University of Cambridge)
Nick Bryan-Kinns (University of the Arts London)
Mercedes Bunz (King's College London)
Baptiste Caramiaux (Sorbonne University)
Sougwen Chung (Artist)
Martin Disley (University of Edinburgh)
Yali Du (King's College London)
Edgar A. Duéñez-Guzman (Co-founder Gibran, ex-DeepMind)
Evelyn Gius (Technical Darmstadt University)
Francisco Gómez Medina (The Alan Turing Institute)
Lauren Goodlad (Rutgers University)
Naama Ilany-Tzur (Carnegie Mellon University)
Leif Isaksen (University of Exeter)
Marina Jirotka (University of Oxford)
Helen Kennedy (University of Nottingham)
David Leslie (The Alan Turing Institute)
Dalaki Livingstone (University of Utah)
Hoyt Long (University of Chicago)
Meredith Martin (Princeton University)
Johanna Nalau (Griffith University, Australia)
Chris Nathan (The Alan Turing Institute)
Ashley Noel-Hirst (University of Edinburgh)
Kirsten Ostherr (Rice University)
Andrew Prahl (Nanyang Technological University Singapore)
Omer Rana (Cardiff University)
Matt Ratto (University of Toronto)
Tobias Revell (Arup)
Jenny Rhee (Virginia Commonwealth University)
Isaac Rutenberg (Center for International Forestry Research and World Agroforestry & Strathmore University Kenya)
Brent Seales (Schmidt Science & University of Kentucky)
Stephanie Sherman (Antikythera)
Richard Jean So (McGill University)
Adam Sobey (The Alan Turing Institute & University of Southampton)
Jack Stilgoe (University College London)
Marion Thain (University of Edinburgh)
Elaine Ubalijoro (Center for International Forestry Research and World Agroforestry)
Ted Underwood (University of Illinois)
Aditya Vashishta (Cornell University)
Matjaz Vidmar (University of Edinburgh)
Matthew Wilkens (Cornell University)
Youyou Wu (University College London)
Sizhe Yuen (The Alan Turing Institute)
Martin Zeilinger (Abertay University)

Workshop participants & technical reviewers

Thanks to **Giada Alessandroni** (EPSRC-UKRI), **Hannah Andrews** (British Council), **Michael Ball** (MRC-UKRI), **Courtney Bates** (University of Edinburgh), **Kaspar Beelen** (The Alan Turing Institute & University of London), **Eamonn Bell** (Durham University), **Emmanouil Benetos** (Queen Mary University of London), **Julia Black** (The British Academy), **Jennifer Cearns** (University College London), **Allaine Cerwonka** (The Alan Turing Institute), **Alan Chamberlain** (University of Nottingham), **Wendy Hui Kyong Chun** (Simon Fraser University), **Shauna Concannnon** (Durham University), **James Dobson** (Dartmouth University), **Eleanor Drage** (Leverhulme Center for Future of Intelligence), **Kerry Francksen** (Centre for Dance Research Coventry University), **Wesley Goatley** (University Arts London), **Jennifer Gold** (ESRC-UKRI), **Jonathan Gray** (King's College London), **Dawn Greenburg** (AHRC-UKRI), **Muntasir Hashim** (Lloyd's Register Foundation), **Katy Henderson** (The Alan Turing Institute), **Daniela Hensen** (BBSRC-UKRI), **Laura Herman** (Adobe), **Ryan Heuser** (University of Cambridge), **Sarah Immel** (University of Edinburgh), **Leo Impett** (University of Cambridge), **Baindu Kallon** (Independent Social Research Foundation), **Esra Kasapoglu** (Innovate UK-UKRI), **Sebastian Laurent-Powers** (AHRC-UKRI), **Emily Lanham** (University of Oxford), **Susan Lechelt** (University of Edinburgh), **Sang Leigh** (Cornell University), **Alex Mankoo** (The British Academy), **Georgia Meyer** (London School of Economics), **Daniela MihAI** (University of Southampton), **Chris Newfield** (Independent Social Research Foundation), **Deven Parker** (University of Glasgow), **Yipeng Qin** (Cardiff University), **Emily Robinson** (University of Exeter), **Jessica Ratcliff** (Cornell University), **Karina Rodriguez Echavarria** (University of Brighton), **Mark Sandler** (Queen Mary University), **Thea Sommerschield** (University of Nottingham), **Allan Sudlow** (AHRC-UKRI), **Christopher Thomas** (The Alan Turing Institute), **Milena Tsvetkova** (London School of Economics), **Kay Watson** (Serpentine Galleries), **Daniel Wilson** (The Alan Turing Institute), **Zheng Yuan** (King's College London)

Executive summary

Artificial Intelligence is rapidly becoming global infrastructure – shaping decisions in healthcare, education, industry, and everyday life. Yet current AI systems face a fundamental limitation: they are shaped by narrow operational metrics that fail to reflect the diversity, ambiguity, and richness of human experience.

This white paper presents a research vision that positions interpretive depth as essential to building AI systems capable of engaging meaningfully with cultural complexity – while recognising that no technical solution alone can resolve the challenges these systems face in diverse human contexts.

We identify a foundational gap:

AI systems increasingly produce and act upon cultural outputs – language, images, narratives – yet lack frameworks for interpreting the cultural content they generate and encounter.

At the same time, AI is entering domains where success is harder to define: areas without clear ground truth that demand contextual reasoning and interpretive judgement. In such cases, traditional benchmarking breaks down.

These interpretive challenges fall precisely within the expertise of the humanities, arts, and qualitative social sciences – disciplines that specialise in understanding cultural meaning, contextual nuance, and interpretive complexity.

This gap creates measurable deployment failures and ethical risks across diverse contexts, limiting AI's effectiveness and global applicability. We've seen this before: early social media platforms were released with minimal contextual safeguards and benchmarked on simplistic engagement metrics – leading to unanticipated societal harms. AI, now entering even more sensitive domains, must not follow the same path.

But the opportunity is greater than the problem: integrating interpretive capabilities could unlock significant advances in AI's ability to solve complex, real-world challenges while ensuring these technologies amplify rather than erode human potential. This is a critical moment to shape AI's foundations – early design choices will steer its trajectory for years to come.

Three critical challenges

The qualitative turn: AI is no longer limited to structured prediction or optimisation – it now operates in tasks that require contextual judgement, cultural nuance, and interpretive reasoning.

The homogenisation problem: The dominance of a few AI architectures propagates design limitations across countless applications and can entrench social inequalities by reinforcing narrow models of reasoning and representation.

The transformation of human cognition: As we engage with complex, interconnected systems of artificial and human agents, AI is reshaping human thinking and work in ways that risk diminishing rather than enhancing human agency and capabilities.

A new research agenda

Doing AI Differently calls for a fundamental shift in AI development – one that positions the humanities, arts, and qualitative social sciences as integral, rather than supplemental, to technical innovation. This creates *Interpretive AI* – systems designed to handle plurality, ambiguity, and contextual meaning as core capabilities.

The core innovations we envision

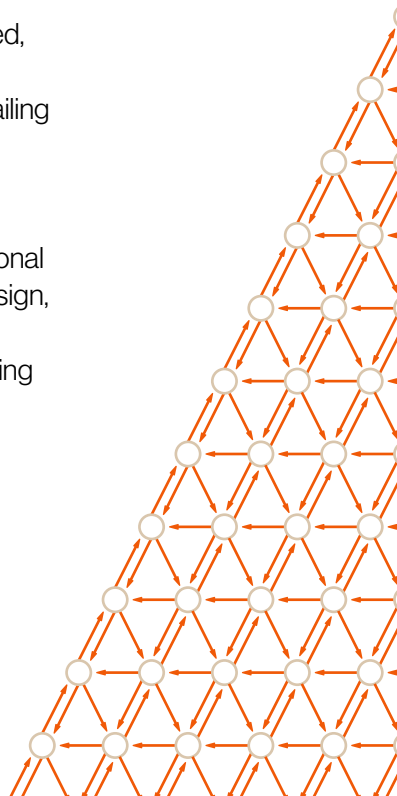
Interpretive technologies: AI systems that represent multiple valid perspectives rather than producing monolithic outputs, enabling more nuanced, culturally sensitive reasoning across diverse contexts.

Alternative architectures for AI: Expanding the AI design space beyond current homogeneous approaches through diverse reasoning paradigms grounded in heterogeneous cognitive, cultural, and planetary processes.

Human-AI ensembles: Developing frameworks for sophisticated, collaborative human-AI systems that strengthen our collective intelligence and enhance rather than replace human capabilities in complex decision-making.

Interpretive AI refers to systems designed to engage with ambiguity, context, and plurality as core capabilities. While this work does not aim to substitute for human interpretive agency, it proposes that humanistic methods – including narrative analysis, cultural reasoning, and contextual sensitivity – can inform how AI systems are designed, trained, and evaluated. We recognise this as a high-risk research direction, precisely because many elements of interpretation resist formalisation – and because it challenges prevailing assumptions in both technical and humanistic domains.

This approach builds on and bridges critical gaps between existing fields. While responsible AI addresses ethical deployment, digital humanities leverages computational tools for cultural research, and human-computer interaction emphasises interface design, this work integrates interpretive reasoning into AI's foundational architecture. Recent breakthroughs like DeepSeek – which achieved competitive performance by integrating humanities scholars directly into technical development teams – demonstrate the measurable value of this approach at scale.



Demonstrating the gap: societal challenges

Sustainability: Scientific consensus on climate change is clear, yet AI-driven pathways for action often ignore the local, cultural, and political specificities that shape real implementation. Interpretive capacity is essential for bridging global models with grounded, diverse realities.

Healthcare: Patients' lived experiences are rich, sensory, and emotional – but often flattened by data-driven systems. Interpretive AI can preserve narrative complexity, supporting better diagnoses, trust, and care outcomes.

Engineering design: Engineering design requires AI systems that can interpret cultural contexts and user meaning, supporting collaborative teams rather than automated optimisation.

Strategic context and momentum

Doing AI Differently has gained significant international traction: 50+ authors and 150+ active researchers across 6 continents, validation by 70+ leading experts, adoption as a UKRI programme theme, and £1M investment by UK's AHRC and Canada's SSHRC for UK–Canada–US collaborations.

This initiative builds on prior work by AI artists, curators, and creative technologists, whose early engagement with AI systems as cultural and interpretive media helped shape the co-creative and epistemic orientation that defines this agenda.

The timing is critical: the systematic erosion of humanities funding is occurring precisely when interpretive expertise becomes technically essential for AI development. The cost of inaction is significant: continued deployment failures, diminished capacity to shape emerging AI economies, and missed opportunities to lead in next-generation AI development.

Roadmap and engagement pathways

Early participants will help shape both the research agenda and the policy frameworks that will define AI's next decade. This white paper provides a concrete roadmap for doing so:

- The research vision and five strategic workstreams to catalyse breakthrough research across disciplines and sectors, outlined in detail in the report.
- A parallel set of policy-facing recommendations, presented in a separate policy note.

Technical foundations will be established by 2026, with demonstrable impact visible by 2030 through AI systems that can effectively operate across diverse cultural contexts.

Call to action

This is more than a report – it is a call to action and a plan for change. We invite researchers, institutions, and funders to join us in this crucial endeavour to unite the humanities, data science, and engineering in shaping the future of AI. This includes deepening collaboration with those whose cultural lives, environments, or rights – and the more-than-human systems they are entangled with – may be shaped by AI, but who are often left out of its design.

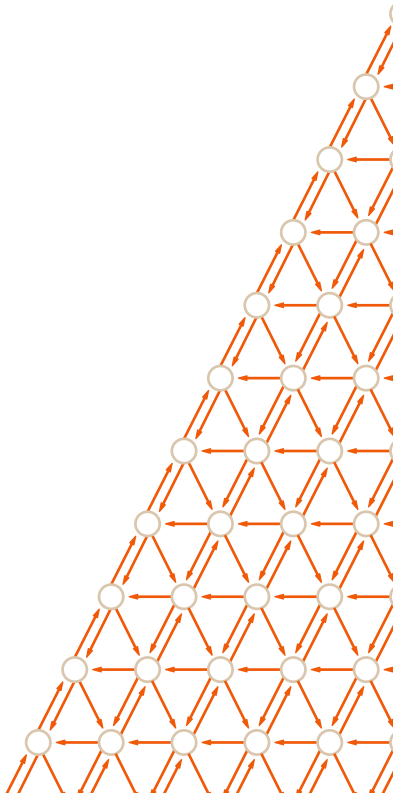
By acting now, funders, institutions, governments, industry, and researchers can help shape AI's trajectory – ensuring that this generational technology enhances human capabilities, reflects global diversity, and delivers positive societal outcomes at scale.

A view from the arts

> ***Abstractions like 'the Machine' don't arrive from nowhere*** *– they're constructed through our existing histories, philosophies, and cultural perspectives. It's easy to forget that* ***there's no such thing as a single artificial intelligence****, because there's no such thing as a single natural intelligence. There's meaning in the data – but* ***it's not the meaning we are given; it's the meaning we make****.*
>
> *Sougwen Chung*
> *(Artist)*

A view from industry

> ***AI needs to be done differently.*** *In its current design as a consumer product, AI has the power to dramatically reduce human agency. Current AI systems take control away from the people who use these tools in important ways: the companies that own these technologies can simply dictate how they function through system prompts and architectural choices. For example, we see the beginnings of AI personalisation designed to enhance corporate profits rather than empower consumers.* ***In my own response to this problem, I left my role at a major tech company to build AI tools based on different principles – to maximise individual human autonomy.*** *This white paper identifies the same core problem from another perspective.* ***Addressing homogenisation and ensuring AI complements, rather than displaces, human capabilities and agency is imperative.*** *Tackling these urgent issues will require coordinated action from academia, industry, policy, and more.* ***This paper offers a crucial vision for how a broad coalition can drive this effort forward.***
>
> *Edgar A. Duéñez-Guzman*
> *(Co-founder Gibran,*
> *ex-DeepMind)*



Preface

Artificial Intelligence (AI) stands at a critical juncture in its relationship to society and more broadly humanity.

As part of UK Research and Innovation (UKRI), the Arts and Humanities Research Council (AHRC) is leading on a range of research and innovation programmes focussing on the ethical and responsible development and deployment of new AI enabled technologies. This supports work that is firmly pro-innovation but equally prioritises a human and planet centric approach for world-leading research and its impacts.

As part of that research ecosystem, the Doing AI Differently initiative has convened and catalysed in a remarkably short time a vibrant international community of researchers and practitioners across six continents. My specific thanks to Professor Drew Hemment and his team who have led on this, along with our partners at The Alan Turing Institute and the University of Edinburgh. This initiative has engaged an expert community around a new and compelling vision of doing AI differently as set out in this White Paper.

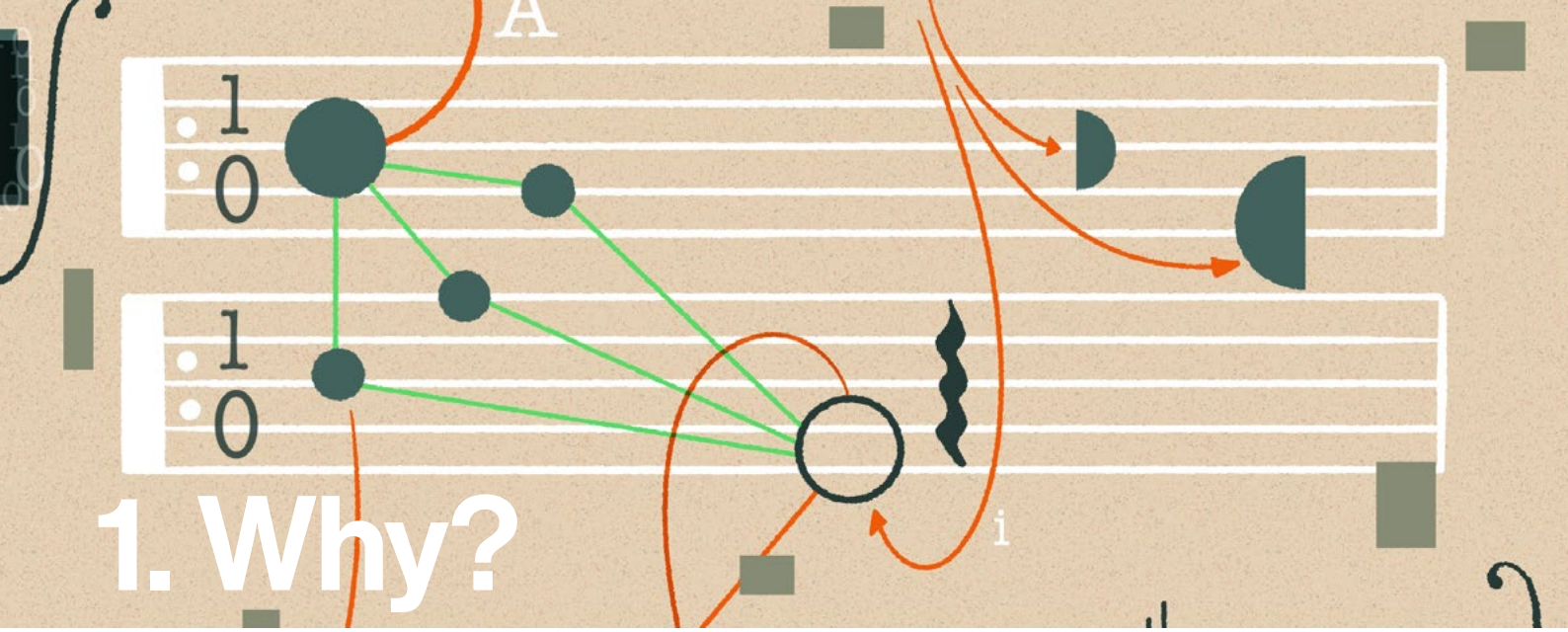
Herein you can read about approaches to deeper interpretive AI capabilities capturing multiple perspectives and semantic depth. Explore new concepts and approaches to AI design that will enable more pluralistic and culturally adaptive AI systems. And move beyond deployment of AI for substitution or assistance models, to foster human-AI technology compacts to achieve outcomes neither could accomplish alone.

This is a timely intervention. The UK Government has set out its ambitions for AI across sectors, research, the economy and society in the AI Opportunities Action Plan. This plan clearly states a need for AI development and deployment that drives economic growth and improved public services, directly benefits how citizens interact with their government, supports world-leading research and innovation, and opens new avenues for human creativity and skills development.

Doing AI Differently is a fundamental component of achieving this. That's why AHRC-UKRI through the UK International Science Partnerships Fund and our partners at the Canadian Social Sciences and Humanities Research Council will be collectively investing a further £1 million to support international research sandpits. These sandpits, under the inclusive direction of Professor Hemment and our partners, will provide a testing ground for the concepts and vision set out in this paper. Our longer-term aim is development of new humanistic approaches in AI data science and engineering towards real-world benefits and impacts.

AHRC-UKRI warmly welcomes further partnerships with UK and international funders to build a global community around this vital work, ensuring the vision set out here can be realised at scale and sustained through collective effort.

Allan Sudlow
Director of Partnerships, Arts & Humanities Research Council (AHRC-UKRI)



1. Why?

The urgent need for this work

1.1 Introduction

> It feels like we are at a threshold moment where, if these reduced notions of intelligence are allowed to stand in for human intelligence itself, then more complex ideas will fade into the background.
>
> *Hoyt Long*
> *(University of Chicago)*

The current wave of Artificial Intelligence – from large language models to agentic systems – underscores a critical need for insights from the humanities, arts, and qualitative social sciences to shape AI's future.

This paper identifies a pivotal inflection point: the qualitative turn, a shift toward systems whose inputs and outputs are cultural. It responds to the growing homogenisation of AI – in both outcomes and design – where systems often prioritise efficiency through general-purpose, single-user models. This narrowing risks constraining AI's representational capacity and undermining its potential for genuine human-AI synergies. These constraints present an opportunity to advance AI's representational scope, architectural diversity, and collaborative potential.

Humanist scholars have already made vital contributions through analysis, datasets, and ethical frameworks. This white paper proposes extending this engagement upstream: integrating humanities perspectives directly into AI's foundational design, architecture, and implementation. Rather than treating humanities insights as external critique, we explore how interpretive approaches from these disciplines might inform computational frameworks while preserving the integrity of humanistic inquiry. This includes expanding beyond anthropocentric assumptions to engage interpretive traditions that foreground ecological interdependence and more-than-human relations.

The central research challenge is to explore whether and how interpretive methodologies can inform AI development, recognising that some aspects of cultural meaning-making may resist or require protection from computational treatment. This approach seeks to strengthen both the contextual sophistication and ethical integrity of AI systems – enabling them to engage with diverse human values by design, not as an afterthought – and to open new pathways toward more collaborative, context-aware, and socially responsive technologies.

1.2 Transformative shifts in AI: challenges and opportunities

We identify three primary challenges which, if addressed constructively, present significant opportunities to both advance the technical foundations and transform the societal impact of AI:

1.2.1 The qualitative turn

The field of AI has undergone a profound transformation in recent years. For most of its history, AI was primarily understood through mathematical and algorithmic processes; its systems were best analysed through the numerical operations that governed their behaviour. However, with the rise of large language models (LLMs), AI has experienced what we identify as a qualitative turn – a shift toward systems whose inputs and outputs are deeply rooted in human cultural and social contexts.

Today's AI systems, particularly LLMs, are fundamentally different from their predecessors. While traditional AI systems often produced numerical outputs or operated within narrowly defined domains, modern AI generates outputs – text, images, and multimedia – that are embedded in human cultural contexts. These outputs don't just resemble human cultural artifacts; they are derived from and respond to the vast corpus of human knowledge, communication, and creative expression. As AI moves into domains where interpretive judgement is required and outcomes are not easily benchmarked, traditional forms of evaluation begin to break down – contributing to the implementation gap between AI's technical performance and its real-world relevance.

Why is this qualitative turn significant? For several key reasons:

1. **Cultural data as foundation:** Contemporary AI models work by ingesting massive amounts of the human cultural record. They are, in essence, bottom-up models of human culture, being steered less by traditional algorithmic design and more by the patterns present in the data they've been trained on.
2. **Socially contextualised outputs**: The outputs of these systems have more in common with documents studied by humanists than equations studied by mathematicians. While the underlying mathematical properties remain important (and work on mechanistic interpretability continues to advance our technical understanding), these properties alone cannot provide a complete account of how these systems function in real-world contexts.
3. **Need for interpretive methodologies**: This evolution demands complementary approaches for analysing and interpreting these systems. The humanities offer essential methodologies – including critical analysis, close reading, and contextual interpretation – for understanding the rich, qualitative outputs of AI and how they interact with human users across diverse sociocultural settings.

The implications of this qualitative turn extend beyond how we analyse AI systems – they fundamentally change how we should design them. Without incorporating humanities perspectives into the core development of AI, we risk creating systems whose outputs are technically sophisticated but culturally impoverished.

This shift is already being advanced through interdisciplinary practices that engage AI not only as a technical system but also as a medium for cultural and interpretive inquiry.

This represents both an opportunity and an imperative: the opportunity to create more nuanced, contextually aware AI systems, and the imperative to ensure that diverse perspectives from the humanities, arts, and qualitative social sciences inform AI development from the ground up – not merely as post-hoc analysis or interpretation of outputs, but as fundamental input to system architecture and design.

1.2.2 The homogenisation problem

A striking feature of today's AI landscape is the notable concentration of effort and investment around a small number of architectures – primarily deep neural networks and reinforcement learning. While other approaches persist and have active research communities, large-scale deployment and funding have converged around a narrow band of scalable methods. This convergence has delivered impressive results but is now associated with stagnating model performance, as similar architectures trained on similar data reach diminishing returns. These limitations are widely recognised within industry and academia.

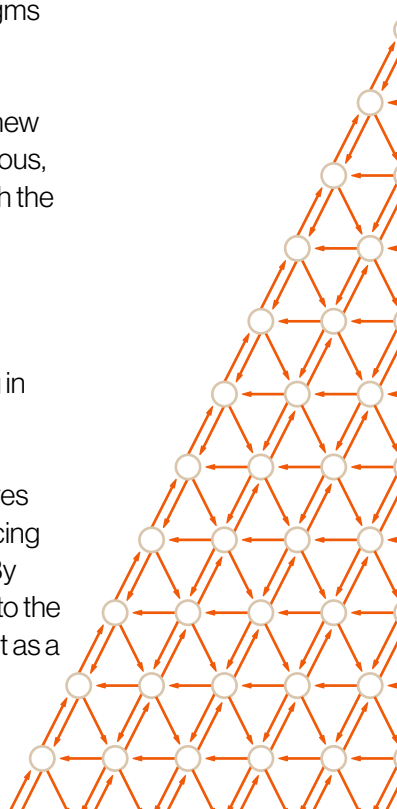
This technical convergence creates several interconnected challenges that must be addressed if AI is to benefit diverse human communities, enhance the full spectrum of human capabilities, and support the ecologies they inhabit. Homogenisation operates not just at the architectural level, but in data pipelines, benchmark practices, and in the underlying assumptions about what constitutes "intelligence" in computational systems. Certain model architectures inherently support some applications and outcomes while excluding others – with the risk of reinforcing existing social inequalities and narrowing the representational space.

Why is this homogenisation problematic? For several key reasons:

1. **Systemic performance stagnation**: Despite growing investment, new models often deliver only marginal improvements. When many systems draw on the same datasets and architectures, performance converges – and further gains are harder to achieve. Innovation may depend on expanding the design space itself.
2. **Narrowed conceptions of intelligence**: We've often allowed "intelligence" to be defined in ways that prioritise easily measured and quantified tasks. We stand at a threshold moment where, if these reduced notions of intelligence are allowed to stand in for human intelligence itself, then more complex understandings may fade into the background. Diverse disciplines can restore complexity and dimensionality to our understanding of intelligence and cognition.
3. **Systemic amplification of limitations**: When similar models trained on similar data are deployed across various domains, their shared blindspots and assumptions become systemic features of our technological landscape. What begins as a technical limitation can manifest as a social and cultural constraint, affecting how AI systems interact with diverse human communities.
4. **Overlooked design alternatives**: The focus on scaling existing approaches has diverted attention and resources from exploring fundamentally different computational paradigms – including neuro-symbolic, embodied, or narratively structured approaches. Here, humanities perspectives can inform not only how AI systems are understood, but how they are designed, developed, and evaluated – and may, in time, contribute to entirely new design paradigms. Various scholarly traditions – including feminist, ecological, indigenous, disability, and postcolonial perspectives – offer distinct analytical lenses that can enrich the development of alternative methods.

It's important to note that identifying patterns of homogeneity can sometimes be analytically useful – for example, when simplifying complex data to reveal trends. But this differs fundamentally from the uncritical reproduction of homogeneous reasoning in systems designed for widespread application.

Addressing homogenisation requires more than simply adding humanities perspectives to existing AI pipelines. These perspectives must inform system design itself – influencing not only how we evaluate AI, but how we conceptualise its function and architecture. By fostering a more plural ecosystem of approaches, we can build systems better suited to the diversity of human needs, values, and contexts – not merely as a matter of fairness, but as a precondition for making AI more useful, robust, and reflective of human potential.



1.2.3 The transformation of human cognition

As people increasingly engage with AI systems, the nature of human thought, agency, and social interaction is being fundamentally transformed. AI is not merely a tool; it is a partner in our cognitive processes, shaping how we access information, make decisions, and understand ourselves and our relationships to others. This emergence of human-AI hybridity extends beyond simple augmentation – it represents a profound shift in what it means to think, create, and participate in culture.

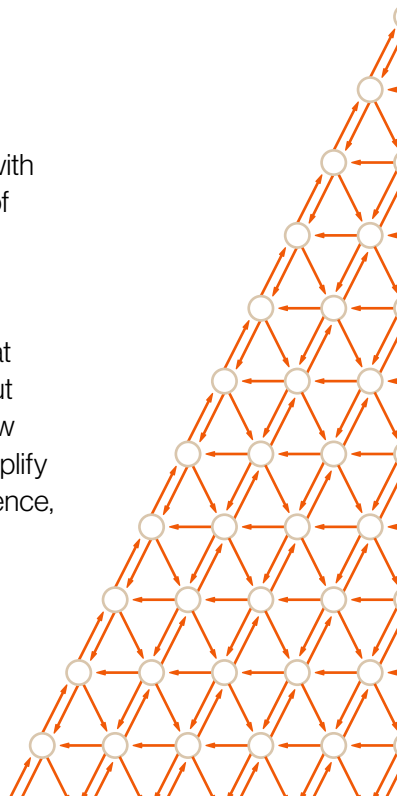
The development of AI has, to this point, been largely organised around the assistant paradigm, with systems performing like search engines, clerks, copywriters, and illustrators. For these uses, fluency and fidelity have been prioritised: systems produce outputs resembling those of competent humans that conform to supplied specifications. But as AI becomes more deeply integrated into cognitive and creative processes, we must look beyond this assistant model toward more complex forms of human-AI collaboration and co-evolution. Human-AI teaming points in this direction, but requires rethinking through the lens of shared agency and cultural context.

Why does this transformation demand humanistic intervention? For several crucial reasons:

1. **Beyond fluency and fidelity**: A genuinely mixed human-AI culture will require systems that can perform as if they are limited in the ways that culturally situated humans are limited – with distinct combinations of perspectives, experiences, and blindspots that drive cultural development. Humanists are uniquely positioned to identify and analyse these varieties of cultural constraint and context that differentiate one community from another.
2. **Hybrid cognitive capacities**: As AI becomes integrated into our cognitive processes, new forms of thinking and problem-solving emerge – neither purely human nor purely artificial, but hybrid. Understanding these emerging cognitive patterns requires frameworks that go beyond technical performance metrics to consider how meaning, agency, and understanding operate within human-AI ensembles. The humanities offer rich traditions for analysing such emergent phenomena.
3. **Ecological and more-than-human perspectives**: AI systems exist within complex webs of connection that extend beyond human communities. An ecological perspective helps us understand how communities develop relationships to AI that reflect their own epistemologies, ways of being, and priorities – both as users of AI technologies and as part of the infrastructure that enables these technologies. This connects algorithmic processes to material ecologies while centring local knowledge systems.

The potential impacts of these transformations are far-reaching. Without careful consideration of how AI shapes human cognition and capabilities, we risk diminishing the very aspects of human experience that make life meaningful. Systems designed with narrow notions of productivity or efficiency may erode our capacity for certain forms of deep thought, creativity, and social connection.

To navigate this uncharted territory, we must develop frameworks for understanding the mutual shaping of human cognition and AI, and design sociotechnical systems that enhance, rather than diminish, our uniquely human capabilities. This is not simply about protecting what is human from AI encroachment, but about imagining and building new capacities – creating systems that allow humans and AI to collaborate in ways that amplify our collective potential while respecting the distinctive value of human agency, experience, and cultural diversity.



Key definition: interpretive vs interpretable

Interpretive approaches *in the humanities involve analysing cultural artifacts within their broader social, historical, and cultural contexts to uncover meaning, significance, and implications – focusing on how technologies shape and are shaped by human experience and societal structures. This differs from* ***interpretable AI****, which refers to technical approaches that make AI systems' decision-making processes transparent and explainable to humans, often through visualisations or explanations of model mechanics. While interpretable AI aims to reveal how a system works, interpretive approaches seek to understand what a system means in its full human context – both are crucial but complementary perspectives for developing AI that truly serves human needs.*

1.3 Articulating the problem space: a transdisciplinary endeavour

The current AI landscape has made impressive strides in addressing ethical considerations, cultural diversity, and human-centred design. However, these efforts could be enriched through deeper transdisciplinary collaboration. Without such integration, AI systems may operate with definitions of human activity and value that, while well-intentioned, do not fully capture the richness, complexity, and diversity of lived experience. To address this opportunity, this initiative explores how AI development might be guided and constrained by perspectives from disciplinary traditions that specialise in cultural meaning, while maintaining space for humanistic inquiry that resists instrumentalisation.

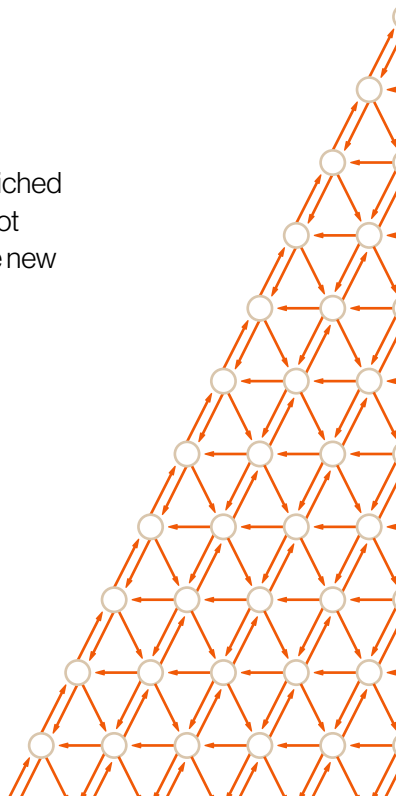
This transdisciplinary endeavour requires recognising the distinct strengths that different fields bring to AI development:

Complementary perspectives on shared challenges:

Complex evaluation frameworks: While computer science has developed sophisticated approaches to evaluating model performance and explaining model decisions, the humanities offer complementary methodologies for understanding outputs in their cultural, historical, and social contexts.

Cultural understanding: Recent technical work on cultural alignment has made significant progress in representing diverse values, but can benefit from humanities approaches that move beyond indexical representation toward deeper contextual understanding of how culture operates.

Human-AI collaboration: Technical approaches to human-AI interaction can be enriched by humanities perspectives on agency, creativity, and meaning-making that explore not just how to make AI serve human needs, but how these collaborations might generate new possibilities beyond current conceptions.



Unique contributions from transdisciplinary engagement:

Productive friction: Humanities perspectives can introduce valuable moments of pause and reflection in development processes, transforming "problems" and "errors" into doorways for deeper understanding and more nuanced systems.

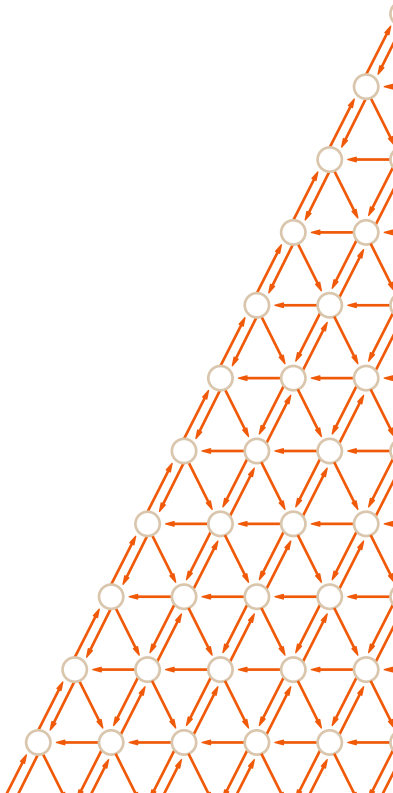
More-than-human frameworks: Work from social sciences, design, and arts exploring "more-than-human" perspectives offers methods and principles for designing systems that consider needs beyond immediate human stakeholders.

Contextual interpretive methods: The humanities bring approaches to interpretation that situate artifacts within broader systems of meaning, helping move beyond technical transparency toward richer understanding of AI's cultural and social implications.

Creative co-creation frameworks: Arts practices including storytelling, speculative design, and artistic inquiry offer methods for shaping the development and evaluation of systems that reflect diverse human imaginaries, rather than replicating existing patterns.

This is not about setting up an opposition between technical and humanistic fields, but rather recognising that the most promising pathways forward emerge through co-creation that integrates these complementary forms of expertise in the early stages of the design pipeline. Neither computer scientists nor humanities scholars alone possess all the tools needed to address the complexities of modern AI development. By bringing together diverse perspectives and methodologies, we can work toward AI systems that better reflect the full spectrum of human and more-than-human needs, values, and possibilities.

> Not every phase of AI development should single-mindedly aim for frictionless utility. Humanities perspectives often prompt developers to slow down and explore why something has gone wrong as much as who, what, or when... These 'problems' are not dead ends but can be doorways to deeper understanding and, ultimately, more nuanced AI systems that better capture the complexity of our world.
>
> *Andrew Prahl*
> *(Nanyang Technological University Singapore)*



In the cybervillage of two hundred billion neurons
human and AI entwined like reeds
planted against the floods.
No Universities, deans, no departments,
only nodes in resonance.

Art no longer fenced from math,
or math from mushroom spores.
We stopped pruning the tree of knowledge;
it bloomed like a weed.

"Design," said Roger, "must obey the wind.
What holds in stillness will fail in a storm."

Irwin taught us: context shifts the blueprint.
Metrics stretch like shadows at dusk.
An AI trained on centuries may fail
when asked the color of a new child's laugh.

Poem created by Roger F. Malina (President, Association Leonardo) using "Fred the Heretic", a custom AI developed by the CyberPoetry team at UT-Dallas and trained solely on the writings of Fred Turner (b. 1941).

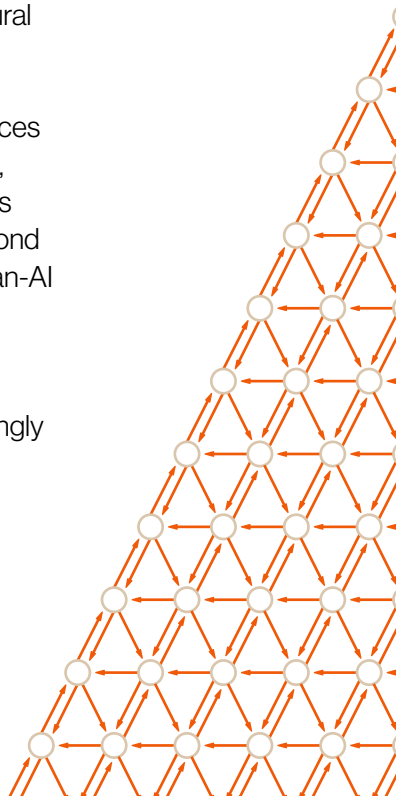
2.What?

The core innovations we envision

This section presents the core research vision at the heart of Doing AI Differently. It outlines three transformative research themes that represent major opportunities for humanities to reshape AI development: **interpretive technologies**, **alternative architectures**, and **human-AI ensembles**. These themes propose new directions that complement and extend current AI approaches, offering ways to engage more fully with ambiguity, cultural diversity, and shared agency.

Each theme defines a distinct area where humanities, arts, and qualitative social sciences perspectives can shape the foundations of AI systems – not only by analysing outputs, but by informing how AI systems are conceived, designed, and deployed. This includes developing AI capable of interpretive understanding, reimagining AI architectures beyond current homogeneous approaches, and creating new paradigms for synergistic human-AI intelligence.

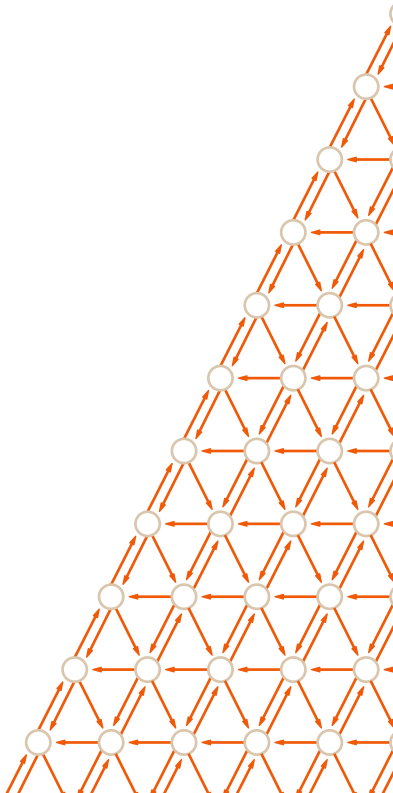
Together, these directions address the fundamental challenge posed by AI's recent qualitative turn, and the shift from numerical to cultural outputs. As AI systems increasingly generate language, images, and other expressive forms, the integration of humanistic insight – via interpretive and ethical reasoning – becomes essential at the level of the technology's core design.



2.1 Objectives

This ambition is supported by five concrete objectives, which guide both the research and the enabling actions proposed in this paper:

1. **Advance interpretive, contextual and perspectival reasoning capabilities** through programmes that integrate humanities, arts, and qualitative social science methodologies into technical development pipelines while safeguarding the critical independence of these disciplines.
2. **Develop pluralistic architectural approaches and evaluation frameworks** that account for the full spectrum of human experience.
3. **Create human-AI ensemble methodologies** that foster collaborative intelligence while enhancing human agency within multi-agent sociotechnical systems.
4. **Demonstrate and secure the vital contributions of deep reflective humanities scholarship** to enable sustained advances in fundamental transdisciplinary understanding on AI.
5. **Build sustainable research capacity and infrastructure** that enables long-term collaboration between humanities scholars and AI developers.



2.2 Interpretive technologies

How can we design and deploy AI systems in a way that captures the inherent ambiguity, context-dependence, and aesthetic dimensions of human meaning?

Contemporary AI systems are often perceived as speaking from a monolithic, "objective" point of view – that of the disembodied model which has seen, read, and synthesised more information than any one human ever could. And yet, even with all that processing power, we know from the humanities that no one such epistemically totalitarian point of view exists in any legitimate sense. The humanities have long recognised that knowledge is situated, interpretive, and inherently multiple. Embedding this understanding into AI architectures supports deeper contextual intelligence – not as external critique, but as a foundation for more socially responsive and value-sensitive systems.

Artistic visualisation by Yutong Liu. Commissioned by Doing AI Differently. Conceptual brief Drew Hemment. CC-BY 4.0

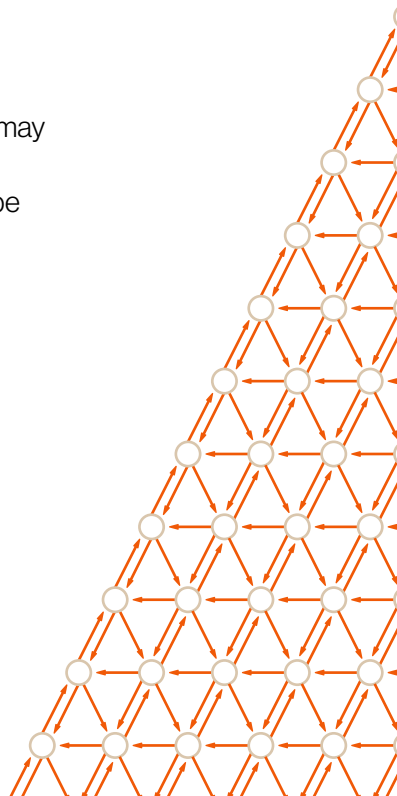
To address these challenges, this research theme will pursue the following objectives:

1. **Develop novel training strategies** that expose LLMs to diverse perspectives and interpretations, enabling them to learn and represent multiple viewpoints on the same topic.
2. **Incorporate contextual information** into LLM representations, allowing them to generate responses that are sensitive to the specific cultural, historical, and social context of the inquiry.
3. **Integrate creative and interpretive methodologies** such as artistic inquiry, speculative design, and embodied creative practices to surface affective, marginalised, and culturally situated ways of knowing – and to shape the epistemic assumptions that guide model and interface design.
4. **Create new evaluation frameworks** that can assess interpretive depth, affective dimensions, and nuance in AI outputs. We need methods and metrics that can capture semantic depth, cultural nuance, and embodied knowledge, while still remaining computationally tractable.
5. **Explore new output formats and modalities** that enable LLMs to express ambiguity, uncertainty, and multiple perspectives more effectively, drawing on insights from creative methodologies to develop novel interfaces and interactive systems.
6. **Develop mechanisms to distinguish AI from human-generated content** based on humanistic frameworks of aesthetic and ethical judgment. This requires a set of heuristics and methods for aesthetic and ethical evaluation translated into computing terms.

This research theme draws upon key theoretical frameworks from the humanities. The humanities and qualitative social sciences have developed sophisticated methodologies for interpreting texts and cultural artifacts that acknowledge the role of context, history, and multiple perspectives. The arts offer methodological approaches that engage with ambiguity through embodiment, affect, and alternative epistemologies. Together, these interpretive and creative practices can inform how AI systems approach complexity in human communication.

The humanities emphasise that knowledge is always situated within particular cultural, historical, and social contexts. This insight can inform how AI systems represent and process information, moving beyond universal claims to acknowledge the specificity of different knowledge traditions.

We recognise that many aspects of human understanding – such as embodied experience, ethical judgement, and relational care – are not only difficult to model, but may be ethically and epistemically irreplaceable. This research explores the boundaries of computational interpretation rather than assuming all cultural meaning can or should be automated.



> Research in rhetoric is very well suited to offer critical contributions to AI research, because it can provide new and different analytical frameworks for interrogating power dynamics and communicative biases/assumptions embedded in language-based systems... Rhetoric skills are now becoming crucially important for all AI users. Importantly, this is not about drawing on insights from rhetoric to create more persuasive AI systems, it's instead about rethinking our approaches to AI development on the basis of critical perspectives that rhetoric has developed over the past decades.

Martin Zeilinger
(Abertay University)

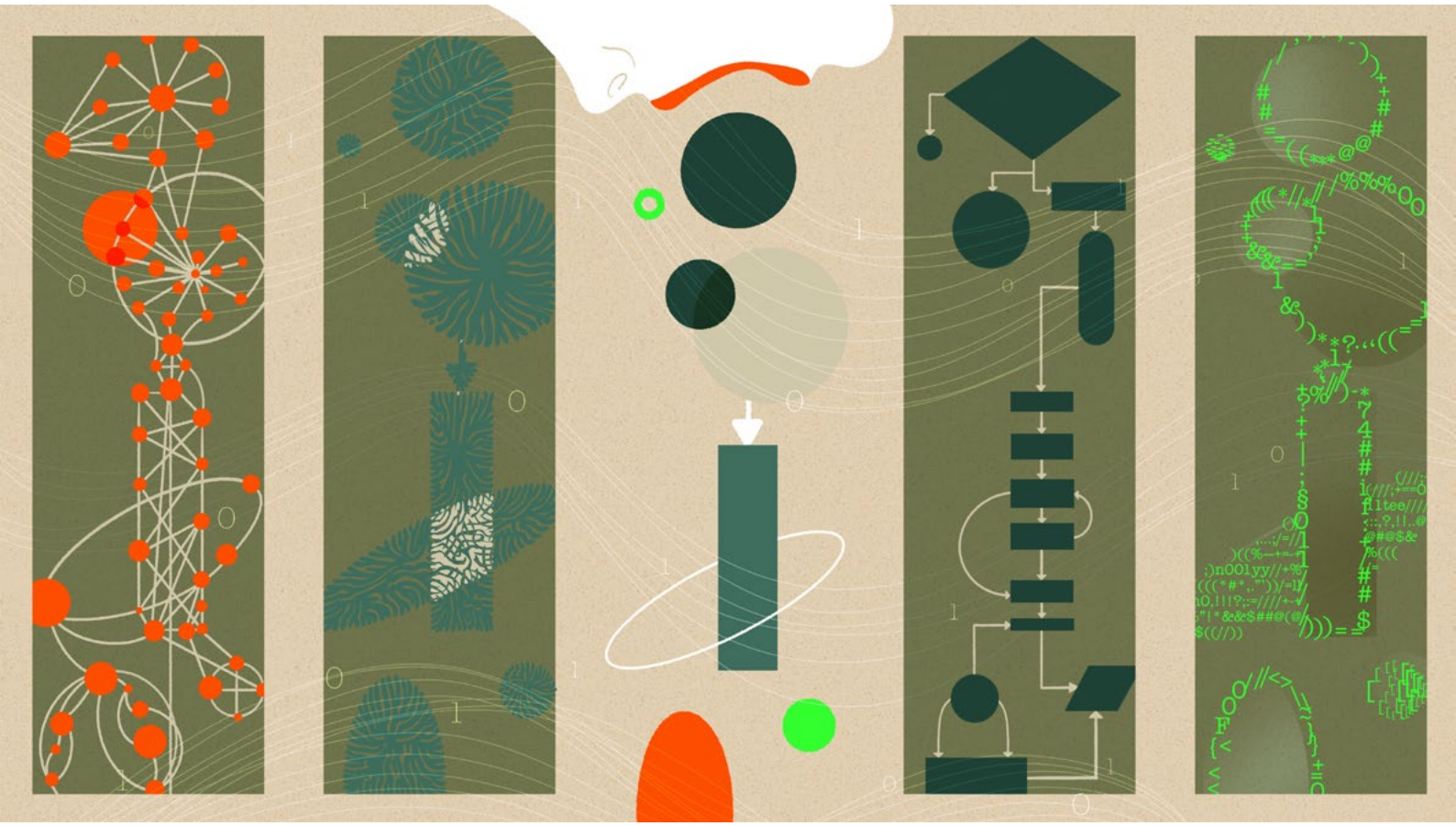
The successful pursuit of this research theme will yield several important outcomes:

1. **New algorithmic approaches** for AI systems that can represent and process multiple interpretive frames.
2. **Enhanced evaluation metrics** frameworks and methods that can assess the interpretive depth and contextual sensitivity of AI outputs.
3. **Innovative user interfaces** that communicate ambiguity and multiple perspectives effectively.
4. **Bespoke data sets** designed to integrate perspectives and approaches from the humanities into AI development.
5. **Prototype systems** that demonstrate the practical value of interpretive approaches to AI.

By advancing this research theme, we aim to create AI systems that enhance rather than flatten the rich interpretive dimensions of human meaning-making, ultimately supporting more nuanced, contextually-aware interaction across human, artificial, and ecological systems.

2.3 Alternative architectures for AI

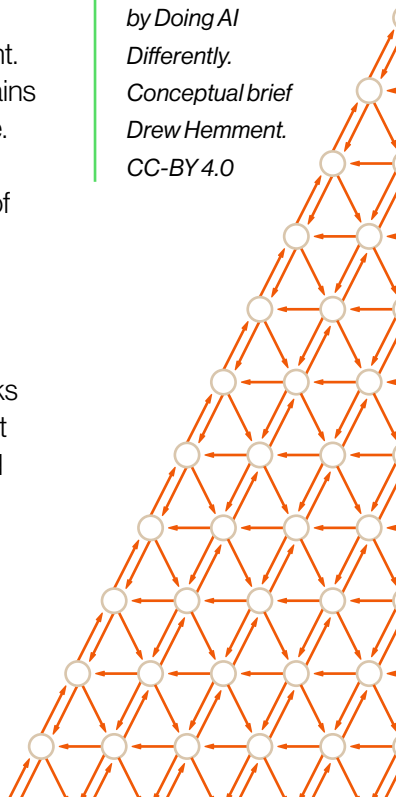
How can we go beyond the homogeneity of current AI architectures, potentially to develop novel neuro-symbolic architectures inspired by humanistic or artistic insight?



Artistic visualisation by Yutong Liu. Commissioned by Doing AI Differently. Conceptual brief Drew Hemment. CC-BY 4.0

The current landscape of AI is dominated by a handful of architectures, primarily deep neural networks and reinforcement learning, that prioritise efficiency and scale. While AI encompasses a range of approaches, these methods receive the majority of investment and deployment, leading to convergence around a narrow band of practices. We use the term 'architecture' here in an expanded sense – not only referring to model topologies, but to system-level design choices spanning data, training, evaluation, and deployment. As these approaches converge on similar data and training pipelines, performance gains are increasingly incremental – signalling a plateau that many in the field now recognise. This homogeneity limits the potential of AI to engage with the complexities of human experience, which is characterised by diversity, ambiguity, and the dynamic interplay of multiple perspectives.

Rather than proposing that humanists take over the engineering process, this theme seeks to create meaningful avenues for humanistic insights to inform and enrich the design space of AI architectures. Specifically, we call for the design of new benchmarks that incentivise engineers to build systems challenging embedded assumptions about homophily, embracing ambiguity, and moving beyond the notion that the future should simply resemble, reinforce, or perpetuate patterns from the past.



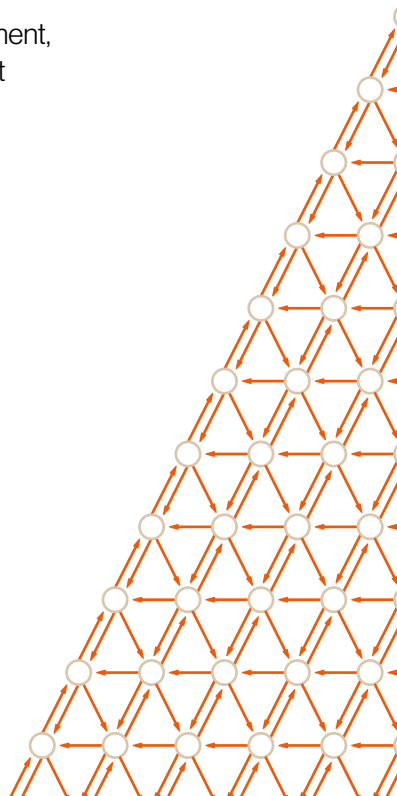
To address these challenges, this research theme will pursue the following objectives:

1. **Identify key characteristics of AI heterogeneity** through critical analysis of current AI benchmarks and their limitations, drawing on humanistic and artistic perspectives to identify dimensions of heterogeneity that are currently neglected.
2. **Develop diverse evaluation approaches** that assess AI systems' ability to embrace ambiguity, engage with multiple perspectives, and adapt to diverse contexts – moving beyond single-metric benchmarks to create incentives for architectural innovation.
3. **Explore alternative learning paradigms** that move beyond loss function minimisation as the primary mechanism for AI development, drawing on humanities and arts insights about how humans learn, adapt, and create through embodied, affective experiences.
4. **Create collaborative frameworks** where humanities scholars, artists, and AI researchers can meaningfully collaborate on architectural innovation, with each bringing their distinct expertise to the process.
5. **Evaluate the impact of new benchmarks** on the diversity of architectures and the range of capabilities supported by AI systems, measuring progress toward more heterogeneous design approaches.

This research theme draws upon theoretical frameworks that challenge the dominant paradigms in AI development, in the full range of AI architecture from pretraining to deployment. The humanities offer rich traditions of engagement with complexity, ambiguity, and context-sensitivity that can inform new approaches to AI architecture. The theme will explore new combinations of methods, informed by humanistic reasoning, that can help address known bottlenecks in current AI performance and generalisation.

These frameworks recognise that current AI systems, while impressive in their capabilities, are constrained by design choices that prioritise certain types of performance over others. The notion of benchmark as used in the AI community is problematic as it tries to rank AI models with single numbers in non-contextualised scales, and the opportunity is to move towards richer and more diverse evaluation of AI models. By bringing humanities perspectives into dialogue with technical development, we can expand the design space and create systems that better reflect the diversity of human cognition and experience.

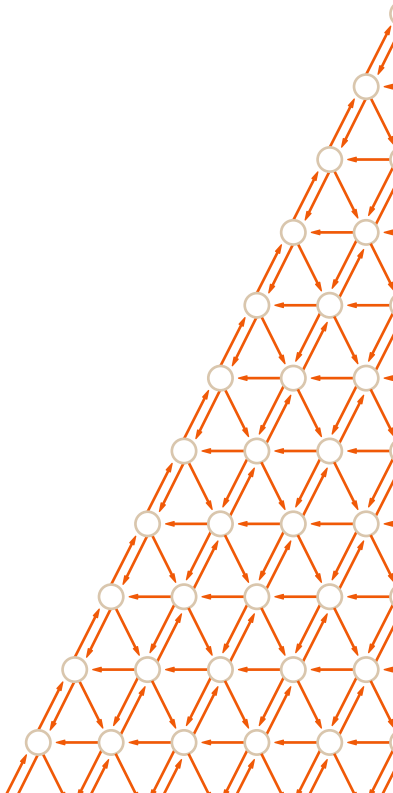
Humanities and arts offer particularly valuable insights into forms of intelligence and creativity that don't fit neatly into current computational paradigms. These include embodied cognition, associative thinking, metaphorical reasoning, aesthetic discernment, and contextual interpretation – all capabilities that are central to human intelligence but difficult to capture in current AI architectures, and which can inform not only interface design but also model structure and architectural logic.



The successful pursuit of this research theme will yield several important outcomes:

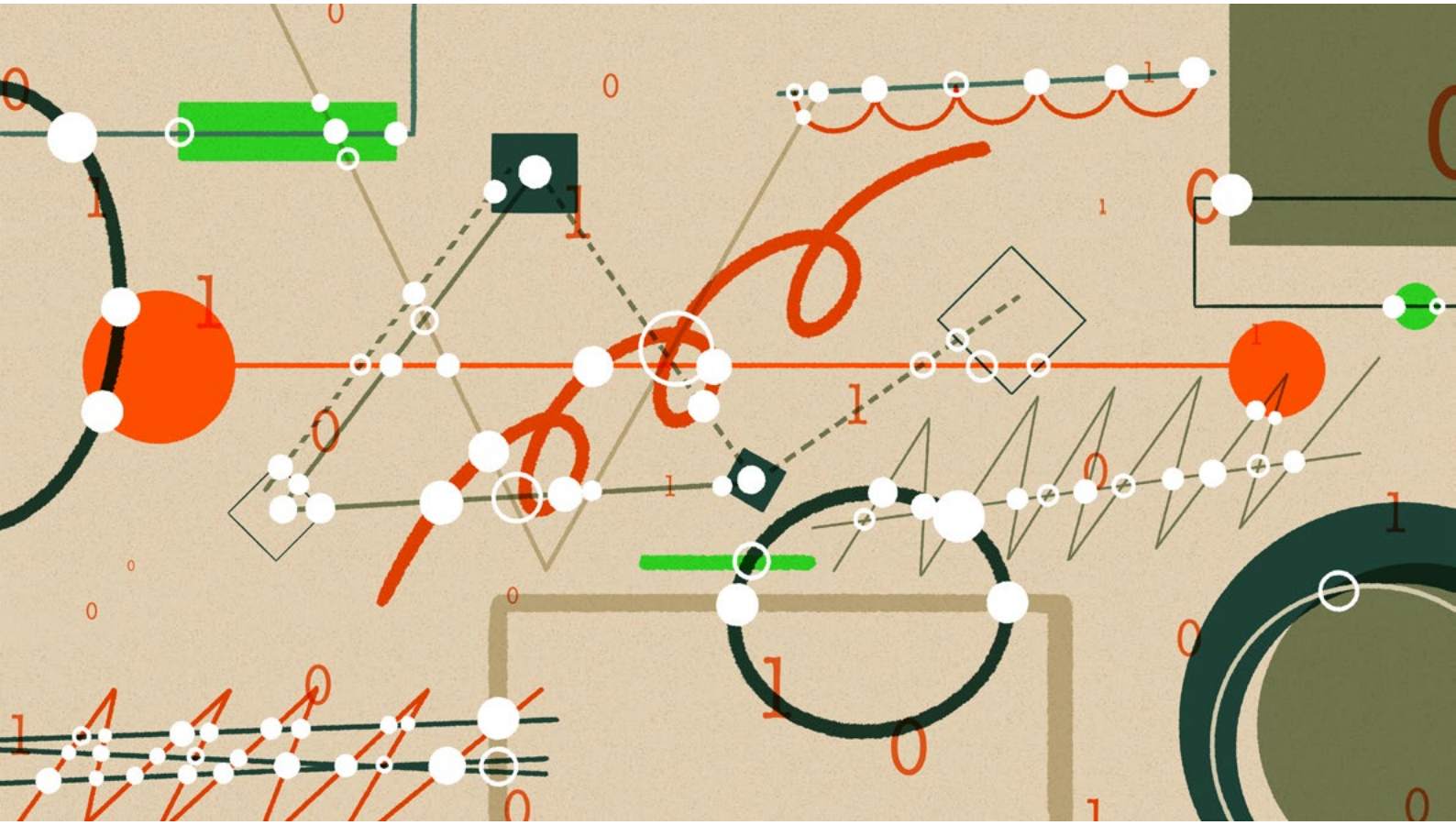
1. **New benchmark datasets and evaluation metrics** that incentivise the development of more heterogeneous AI architectures.
2. **Prototypes of novel combinations of existing and emerging methods**, informed by humanities perspectives, that open up underexplored directions in system design and evaluation.
3. **Frameworks for interdisciplinary collaboration** that enable meaningful participation of humanities scholars in AI architecture development.
4. **Critical analyses of current AI architectures** that identify opportunities for innovation beyond the dominant paradigms.
5. **New theoretical models** that bridge humanities insights and computational implementation.

By advancing this research theme, we aim to expand the design space of AI architecture beyond its current homogeneity, creating systems that better reflect and engage with the richness and diversity of human experience and cognition. By integrating established techniques with insights from humanistic reasoning, this theme may help address known bottlenecks in AI performance – and, over time, open up entirely new directions in system design.



2.4 Human-AI ensembles

How can we build sociotechnical systems that draw on a larger range of interactive configurations – particularly in such a way to ensure that the effect of AI participation is to enhance, rather than replace, human capabilities and ingenuity?



Perspectives on human-AI interaction often draw on a limited set of configurations, typically focused on dyadic relationships between a single human and a single AI system. This narrow framing frequently leads to an assumption of interchangeability: when an AI agent is introduced into a system, a human must come out. Rather than viewing AI as either a replacement for human capabilities or as a simple assistant, this research theme explores the complex networks and arrangements through which humans and AI systems can interact, creating sociotechnical systems that amplify human potential and foster collective intelligence.

Artistic visualisation by Yutong Liu. Commissioned by Doing AI Differently. Conceptual brief Drew Hemment. CC-BY 4.0

To address these challenges, contributors to this initiative have identified the following research directions:

1. **Map the design space of human-AI interaction structures** beyond simple dyadic relationships, exploring a diverse range of configurations through which humans and AI systems can interact within complex sociotechnical networks. This exploration will help identify previously overlooked arrangements with potential for enhancing human capabilities.
2. **Understand information transformation processes** across human-AI boundaries, examining how qualitative human insights and quantitative computational processes can be meaningfully integrated to support decision-making, creativity, and problem-solving that neither could achieve independently.
3. **Develop frameworks for meaningful human agency** within AI-enhanced environments, ensuring that AI systems amplify human capabilities and intentions rather than constraining or replacing them, with particular attention to preserving autonomy and self-determination.
4. **Investigate the social and cultural dimensions** of human-AI ensembles, particularly how they affect power dynamics, community well-being, and equity, with emphasis on developing AI with and for communities rather than imposing technical solutions upon them.
5. **Create principles and methodologies for ethical design and governance** of human-AI ensembles that prioritise human flourishing while acknowledging the complex interdependencies that emerge in these sociotechnical systems.

This research theme draws upon humanistic theories of agency, adaptation, and distributed interaction to reimagine the relationship between humans and AI. Rather than positioning AI as a competitor or substitute for human capabilities, these frameworks explore how AI can serve as a substrate that facilitates and enhances interpersonal interactions and human engagement with their environment and resources.

Viewing human-AI interactions as collaborations can misleadingly place AI systems in a position of comparison with humans. Human-AI ensembles extend beyond conventional human-AI teaming by exploring complex networks and configurations where humans and AI systems interact within broader sociotechnical systems, moving past simple dyadic relationships toward collaborative arrangements that enhance collective intelligence while preserving human agency.

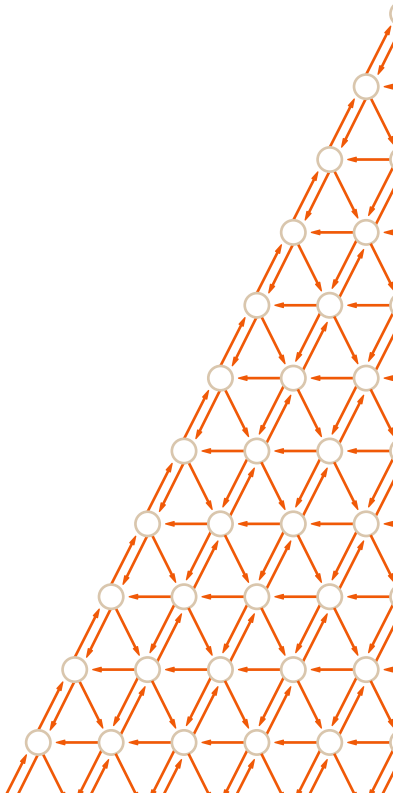
The theme also acknowledges that we currently have a limited understanding of how complex networks of human and AI intelligences will interact in practice. Designing such ensembles requires not only empirical insight, but also critical attention to how structural inequalities, cultural assumptions, and institutional incentives shape these systems and their effects on collective well-being. At the same time, a more-than-human perspective reminds us that these systems operate within broader ecological and sociotechnical networks. Insights from other domains – including how non-human agents such as animals, infrastructures, or institutions influence human behaviour and meaning-making – can help illuminate these complex, interdependent dynamics.

> Building and growing AI is a force that can have devastating and/or monumental impacts on communities, large and small. Through Humanities, we can guide conversations surrounding ethical and equitable approaches towards communities' usage of AI and how it can be addressed holistically in the places we work and live... We need to develop AI with the intent that it is for the community. For the community means it needs to be made with the community. Past, present and future communities cannot be sold solutions but collaboratively build pathways forward.
>
> *Dalaki Livingston (University of Utah)*

The successful pursuit of this research theme will yield several important outcomes:

1. **New models and frameworks** for understanding and designing human-AI ensembles that enhance human capabilities across diverse domains.
2. **Empirical insights** into the dynamics of information flow, decision-making, and emergent capabilities within complex human-AI networks.
3. **Design principles** for sociotechnical systems that prioritise human agency, well-being, and flourishing.
4. **Methodologies for community engagement** in the development and deployment of AI systems, ensuring they serve the needs and values of diverse communities.
5. **Policy recommendations for ethical governance** of human-AI ensembles that address issues of power, equity, and social impact.

By advancing this research theme, we aim to move beyond simplistic substitution models of human-AI interaction toward a nuanced understanding of how humans and AI systems can function together in complex, dynamic ensembles. This approach recognises that the most promising future for AI lies not in replacing human capabilities but in creating sociotechnical systems that amplify and extend human potential while preserving and enriching what makes us uniquely human.



2.5 Building an interdisciplinary community

A vibrant interdisciplinary community has formed around the need for interpretive, context-aware approaches to AI, through purposeful convening at a moment of emerging consensus. Recognising recent developments in AI created latent opportunities for novel disciplinary constellations, the initiative systematically engaged researchers who were already sensing these possibilities. The community now demonstrates strong potential through sustained collaboration, shared conceptual focus, and wide-ranging institutional engagement across six continents.

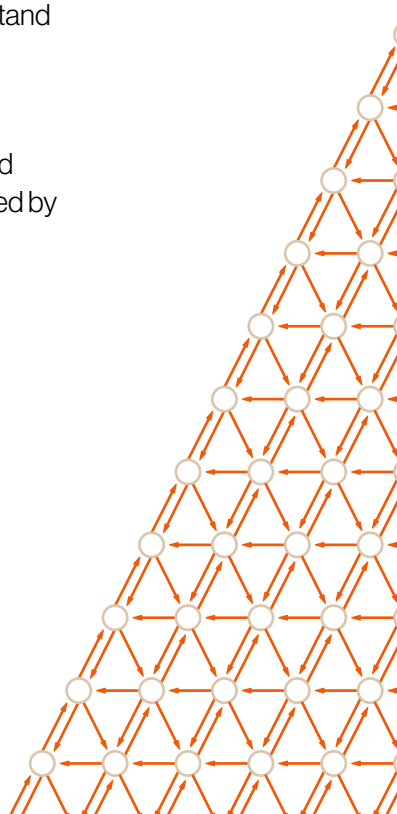
Academic communities often form when existing frameworks prove inadequate for addressing emerging challenges – as seen with cognitive science in the 1960s or digital humanities in the 1990s. The Doing AI Differently initiative identified such a moment and used design-led methodologies to activate alignment between researchers across disciplines who were independently recognising similar needs and opportunities.

Over 18 months, this initiative has systematically engaged 150+ researchers, hosted three international workshops, and produced collaborative outputs including this white paper. Significantly, we have since discovered complementary efforts and communities gravitating in similar directions, confirming the potential for a substantive new research area to emerge. Rapid institutional adoption – including £1M+ in funding and policy integration across multiple countries – demonstrates genuine alignment with strategic research priorities.

This initiative was catalysed by an invitation from the engineering community and built on prior work by AI artists, curators, and creative technologists engaging with interpretive and epistemic questions as co-creators of AI systems. Through collaboration with researchers in responsible AI, digital humanities, human-computer interaction, and data-centric engineering, it crystallised around questions no single field could address alone.

The emerging community is marked by an ethos of collaboration, openness, and deep integration – valuing both technical rigour and interpretive depth, bridging conceptual and applied work, and foregrounding cultural sovereignty, representational justice, and epistemic agency. It views AI development as fundamentally enriched by diverse cultural and humanistic perspectives. Through collaborative research and cross-disciplinary partnerships – including emerging work on *computational hermeneutics* – this community is pioneering approaches that could fundamentally transform how AI systems understand and interact with human meaning.

While this represents significant progress, deeper engagement remains essential – particularly with industry practitioners developing these technologies, researchers and communities across the Global South, and those whose lives and rights may be shaped by AI but remain excluded from its design.



2.6 Strategic positioning

Doing AI Differently identifies an emerging area of research that builds upon and complements the strengths of adjacent domains while seeking to bridge gaps between them. Where digital humanities often applies computational methods to humanities questions, and AI ethics frequently operates at the policy level, Doing AI Differently creates pathways for humanities, arts, and qualitative social science perspectives to directly shape how AI systems are conceived, designed, and built. This approach complements both critical analysis and applied development, aiming to influence AI's fundamental architecture while maintaining the nuance and contextual sensitivity characteristic of humanistic inquiry.

Field	Shared approaches	The extension
Digital humanities	Application of humanistic methods to digital artifacts; focus on interpretation and cultural context	Moves beyond using AI as a tool for humanities research toward directly informing AI architecture and design
Science & technology studies	Analysis of sociotechnical systems; examination of how values shape technology	Combines critical analysis with active participation in technological development rather than primarily focusing on analysis
Critical data studies	Critical examination of power structures in data systems; interpretive methodologies	Extends critique into constructive design intervention, particularly at the architectural level
Human-Computer interaction	Centring human experience; multi-disciplinary approaches; enhancing human capabilities	Emphasises interpretive approaches and fundamental architecture questions alongside interface considerations
Responsible AI / ethics	Integration of ethical considerations; focus on human values	Complements policy-level work with technical implementation of humanistic principles at the architectural level
AI alignment	Integration of human values into technical systems; architectural focus	Enriches technical approaches with deeper contextual understanding and interpretive frameworks
AI arts	Integration of creative practices; embodied evaluation methods; interest in human-AI collaboration	Broadens focus beyond creative applications to fundamental questions of AI architecture and societal systems
Social computing	Focus on social aspects of technology; system design	Strengthens attention to meaning-making, interpretation, and qualitative dimensions alongside quantitative metrics

	AI as tool	Sociotechnical analysis	Critical analysis	Interpretative methods	Interaction design	Foregrounds cultural context	Value integration	Shapes core architecture
Doing AI Differently	●	●	●	●	●	●	●	●
Digital humanities	●	○	●	●	○	●	○	○
STS	○	●	●	●	○	●	○	○
Critical data	○	●	●	●	○	●	○	○
HCI	●	○	○	○	●	○	○	○
Responsible AI	○	●	●	○	○	○	●	○
AI alignment	○	○	○	○	●	○	●	●
AI arts	●	○	●	●	●	●	○	○
Social computing	●	●	○	○	●	○	○	○

2.7 Real world case studies

2.7.1 Sustainability case study

The challenge

Decarbonising at the necessary scale and speed requires more than technical solutions – it demands systems that can navigate complex social, cultural, and political realities. Traditional AI approaches to climate challenges often produce homogenised models that fail to account for diverse contexts, limiting both technological effectiveness and implementation success. This homogeneity creates a fundamental gap: while scientific consensus on climate change exists, the pathways to emissions reduction remain fragmented across different cultural, geographical, and socioeconomic contexts. Having robust AI-powered applications that are applicable to diverse decision-contexts will be critical in scaling up climate action to achieve just decarbonisation. A key challenge is the matching of global frameworks with local context and how to capture local level decision-making and knowledge that support global progress both on reducing emissions and adapting to climate change.

The vision

The research programme described in this white paper describes several routes for addressing this challenge. With the specific goal to decarbonise transport, energy and infrastructure, this work will tackle how AI can effectively engage with the uncertainty, nonlinearity, and contextual dependencies that define real-world decarbonisation challenges. Traditional AI systems struggle with these complex factors. New methods of the kind outlined in this paper are needed to handle heterogeneity at scale while maintaining local relevance – creating systems that can represent and reason across multiple valid frameworks simultaneously.

These innovations enable more effective climate action by:

Contextualising global models with local knowledge: Creating systems that bridge scientific projections with situated understanding of implementation barriers.

Mediating across polarised perspectives: Enabling productive dialogue where ideological divides have previously stalled climate progress.

Identifying context-sensitive interventions: Revealing decarbonisation opportunities that homogeneous approaches overlook.

The applications

Technical innovations of this kind will open new opportunities for low-carbon transition in areas where conventional approaches have stalled. While there is a large range of potential impact, specific examples include the following:

Heat-resilient urban planning: AI-based technologies are beginning to enable cities to develop locally appropriate, low-carbon cooling strategies by integrating community values, cultural practices, and technical specifications. An increasing number of AI-powered applications (e.g., Google's Heat Resilience Tool) use Earth Observations data with thermal imagining to produce fast real-time data and advice for cities and local governments where they should be planting trees/vegetation to reduce urban heat island effects.

Energy system resilience: It is crucial to increase the resilience of infrastructure to climate extremes through adaptable models that incorporate both technical metrics and social factors. Understanding changes in trends and the increasing impact on infrastructure enables the development of new energy system resilience strategies informed by AI. For example, smart sensors are already used to monitor conditions across major infrastructure to assist in making decisions that can avert catastrophic failures during extreme events.

Implementation pathways in complex settings: AI can help overcome entrenched barriers to decarbonisation by revealing context-sensitive approaches in politically challenging environments. Using AI-models can produce new framings that are more acceptable and applicable to broader sets of stakeholders regardless of political leanings.

AI, humanities, and sustainability

*The relationship between AI and sustainability presents two distinct but equally important challenges. On one hand, **sustainable AI** demands we address the environmental footprint of AI systems themselves, where efficiency improvements often paradoxically increase resource consumption (as demonstrated by data centres now consuming up to 20% of energy in regions like Ireland), requiring frameworks that consider broader sociotechnical contexts rather than mere technical optimisation. On the other hand, **AI for sustainability** requires moving beyond incremental improvements to existing systems toward imagining entirely new paradigms for sustainable industries and ways of living, lest we limit ourselves to minor refinements of fundamentally unsustainable systems. Both challenges require integrating humanities perspectives to critically examine assumptions, envision alternative futures, and understand the complex interplay between technology, human behaviour, and planetary systems.*

2.7.2 Healthcare case study

The challenge

Physician and nurse shortages, burnout across the healthcare profession, and rising costs pose major challenges to health systems operations, reducing patient access to high quality, humanistic care. Natural language-based, consumer-facing tools such as ChatGPT seem to offer solutions to these challenges by increasing workforce efficiency and access to virtual care. But at present, LLMs require significant human oversight and review to ensure that errors and biases are not introduced that could render medical records inaccurate and potentially harmful to patients. One issue is that AI is currently based on quantitative reasoning and data, and human experience is made up of many sensory, emotional, and personal factors whose details are often reduced, distorted or lost in the process of transforming personal health narratives into data. The transformative potential of AI for health will be limited, or potentially undermined, if LLMs for health are developed without the involvement of researchers with deep, nuanced understanding of the diverse human experiences of illness and healing that medical humanities can provide.

The vision

This white paper offers a route to bolstering standard AI systems to deal with these complex factors. For example, AI tools such as LLMs will be increasingly used to summarise, interpret, or imitate human speech in healthcare settings through the use of ambient listening systems that transcribe and categorise doctor-patient conversations. These tools risk losing or distorting the complete picture of the patient as a person. While sentiment analysis and facial recognition programmes attempt to capture unspoken dimensions of meaning, these tools are rarely developed with input from the patients who are being interpreted or humanities partners who are skilled at nuanced, contextual interpretation of multimodal representation and communication.

Likewise, similar innovations could target the following objectives:

Developing frameworks for meaningful human agency: Creating systems where patients can provide direction or feedback to ambient listening systems, working collaboratively with AI to increase patient-centred care and bring the patient's voice directly into the medical record.

Investigating the social and cultural dimensions of human-AI ensembles: Creating opportunities for diverse communities to identify the specific social or cultural concerns that make their healthcare needs unique.

Creating principles and methodologies for ethical design and governance: Establishing community-based ethical principles to increase the trustworthiness of both AI and healthcare systems.

The applications

These innovations open new opportunities for human-centred care in areas where conventional approaches have stalled. These include:

Ambient listening systems: Enabling patients to participate directly in the review and annotation of transcriptions from doctor-patient interactions.

Personalised health LLMs: Increased heterogeneity of training data and feedback from patients and doctors will improve the quality of direct-to-patient health information.

Integration with the non-medical aspects of healthcare: Addressing issues related to privacy, linguistic diversity, user variability, and accuracy in high stakes medical environments to improve humanistic care and reduce healthcare professional administrative burden.

Together, these and similar developments could yield demonstrable improvement in measures such as time to diagnosis, patient trust in culturally appropriate care, and physician satisfaction with AI-augmented tools.

2.7.3 Engineering design case study

The challenge

Engineering product designers face a scalability dilemma. They must create increasingly complex products for diverse users while balancing safety, sustainability, and performance across lifecycles. AI and machine-learning tools are already used for tasks such as concept generation and design optimisation. However, their transformative potential remains constrained by a core challenge: how to integrate AI into design teams in ways that enhance, rather than undermine, human expertise, creativity, and contextual judgement.

To date, the use of AI tools in engineering has focused primarily on generating design options. But without attention to how designers think, collaborate, and interpret problems, these systems risk information overload, poor uptake, or a retreat to familiar solutions. What's missing is the ability to embed AI within design processes in ways that support nuanced, interpretive reasoning – an area where humanities-informed approaches can add critical value.

The vision

This white paper envisions human-AI design ensembles that support design teams in navigating complex decision spaces. These ensembles would synthesise multimodal data, surface actionable insights, and reflect the social, cultural, and environmental contexts in which products are used. Instead of presenting options in isolation, such systems would help designers interpret meaning, explore trade-offs, and communicate their decisions transparently.

Two key objectives define this vision:

Developing human-AI design ensembles: Creating systems that help design teams efficiently navigate feature-rich, data-intensive environments while supporting trust, explainability, and creativity.

Generating actionable design narratives: Embedding the diversity, ambiguity, and richness of user experiences and product lifecycles into structured insights that guide innovation.

The applications

These capabilities can bring immediate value to several core areas of engineering design:

Design space exploration: While current AI systems (such as Artificial Neural Networks and Generative Adversarial Networks) can assess aesthetics or form based on geometric and visual features, human-AI ensembles could support designers in understanding the broader significance of design choices – incorporating values like sustainability, accessibility, or brand meaning into early-stage exploration.

Obtaining product insights at scale: By analysing usage data from thousands of products, interpretive AI could uncover latent needs, recurring issues, or unanticipated user behaviours. These insights would go beyond performance metrics to reflect the lived realities of diverse users and contexts.

Context-rich design evaluation: Existing AI tools evaluate innovation or manufacturability in abstract terms. Future systems could provide narrative-based assessments that reflect how products perform over time, across cultures, or under different stakeholder expectations – helping design teams account for ambiguity, risk, and change.

Incorporating interpretive reasoning into engineering AI does not displace technical expertise – it enriches it. These approaches offer a pathway toward design systems that enhance human understanding, bridge knowledge silos, and ultimately lead to safer, more inclusive, and more sustainable products.

2.8 Navigating risks and unintended consequences

Why interpretive AI must be developed with interpretive responsibility

The development of interpretive AI brings powerful new capabilities – but also new risks. As AI systems grow more adept at inferring nuance, context, and cultural meaning, they may enable more sophisticated manipulation, reinforce dominant narratives while marginalising alternatives, or displace human creativity under the guise of efficiency.

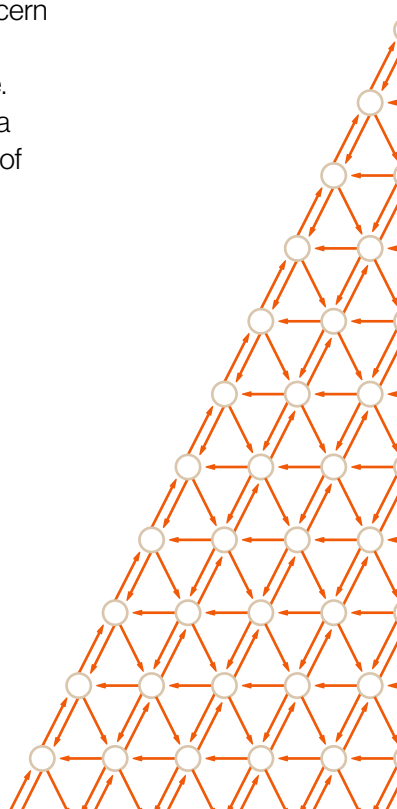
These concerns are particularly acute for creative communities already experiencing displacement through uncompensated appropriation of their work. These risks highlight fundamental questions about the boundaries of computational interpretation. While avoiding interpretive AI may not eliminate dangers, pursuing it requires careful consideration of which forms of cultural meaning-making should remain distinctly human and how to prevent the displacement of essential interpretive capacities.

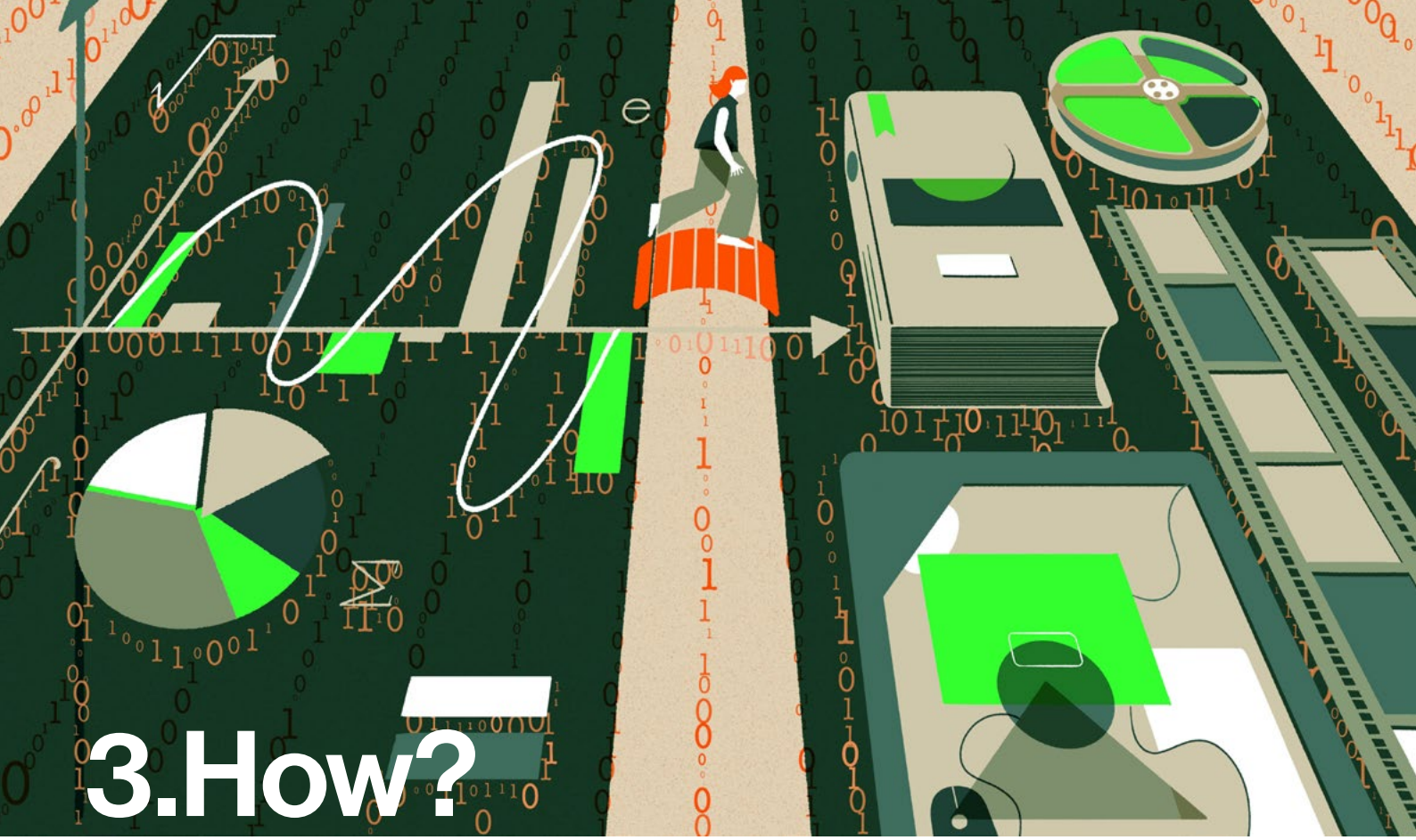
Key areas of concern include:

1. **Sophisticated manipulation**: Culturally targeted persuasion will require new frameworks for identifying and countering manipulation in advertising, political messaging, and platform design.
2. **Cultural appropriation at scale**: Systems trained on cultural contexts may exploit that knowledge without community consent – necessitating community-controlled governance and transparent data practices.
3. **Interpretive homogenisation**: More powerful systems could paradoxically narrow interpretation, reinforcing dominant narratives and sidelining dissent.
4. **Creative and cultural displacement**: As AI enters interpretive domains, it may displace human roles – from translation to artistic practice – calling for models that sustain meaningful human agency.

Meeting these challenges requires co-designing safeguards with humanists, ethicists, technologists, and affected communities – embedding plurality, provenance tracking, and contestability into system architecture from the outset. There needs to be parallel concern for new AI business models and supply chains: with different (fairer, more open, more sustainable) approaches to intellectual property, data-scraping, training and inference. Interpretive responsibility must evolve in parallel with interpretive capability, guided by a shared commitment to human dignity, epistemic justice, and the responsible exercise of power, ensuring these systems serve human and cultural flourishing.

See also: 3.4 Barriers and Risks.





3.How?

The implementation pathway

Realising the vision of Doing AI Differently will require structural and institutional change in how research is supported, and how innovation moves from ideas to adoption. It calls for coordinated action across sectors and disciplines.

This section sets out the strategic workstreams and enabling mechanisms needed to embed humanities, arts, and qualitative social science perspectives as a lasting force in AI development. Alongside the workstreams, we outline integration models, indicative resource frameworks, and success metrics to guide funders, institutions, and research leaders.

Together, these elements are designed to ensure that the activities needed to deliver the Doing AI Differently vision can be pursued effectively and sustained over time.

3.1 Methodology and community engagement

This white paper emerges from an 18-month process of international community engagement designed to identify and articulate research priorities that transcend traditional disciplinary boundaries. Rather than presenting a predetermined agenda, the initiative employed collaborative visioning methods to surface shared challenges and opportunities at the intersection of AI development and humanities scholarship.

***Design-Led Methodology**: Drawing on Open Prototyping – a design-led method for surfacing shared questions and shaping collaborative inquiry across disciplines – the initiative used custom design canvases to articulate "problem spaces": landscapes of questions, resources, and relationships that cannot be addressed within existing disciplinary boundaries. These tools facilitated productive dialogue across different epistemic cultures while maintaining intellectual coherence.*

***Distributed Authorship Model**: The white paper itself represents a novel approach to collaborative academic writing. A core team developed a seed draft, which was then enriched through structured contributions from 50+ scholars, workshop feedback, and iterative refinement. This process ensures the document reflects collective insight rather than institutional perspective.*

***Demonstrated Impact**: The methodology's effectiveness is evidenced by systematic engagement with 150+ researchers across six continents, rapid institutional adoption including £1M+ in UK-Canada funding, and the subsequent emergence of complementary research efforts, indicating genuine alignment with strategic research priorities.*

Full methodology documentation is available at
www.turing.ac.uk/news/publications/doing-ai-differently

3.2 Workstreams

To reflect the priorities raised by contributors across disciplines and sectors, a set of workstreams has been collaboratively formulated. They are intended to guide funders, institutions, and research leaders in supporting the growth of this emerging field.

The five workstreams form an integrated framework: three focus on transformative research directions (W1–W3), while two focus on the enabling infrastructure and institutional pathways needed to realise that transformation (W4–W5).

Together, they provide both the intellectual foundations and the structural support for sustained, cross-sector impact.

Workstream 1: Develop interpretive AI foundations Establish an open ecosystem to foster a plurality of interpretive evaluation frameworks, representation methods, and integration tools – essential for effective deployment in diverse societies and sectors.

Workstream 2: Expand AI design pathways Support interdisciplinary labs to explore new combinations of model design, training, data, and evaluation – informed by humanities insight – to diversify current system development.

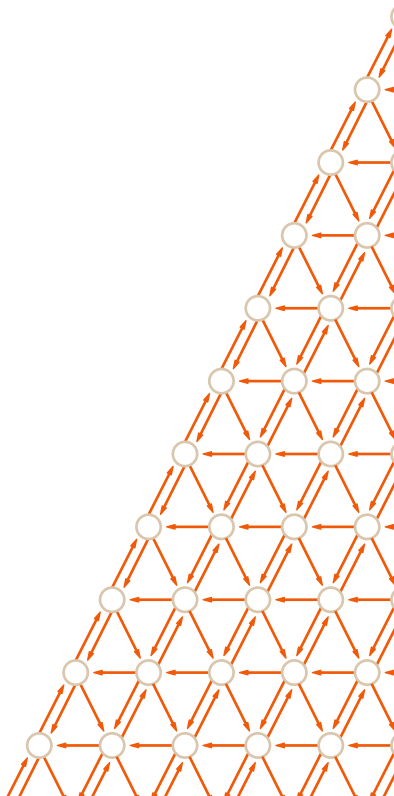
Workstream 3: Enable human-AI ensemble systems Advance research and field experiments in collaborative AI that preserves human agency while enhancing decision quality, with pilots in healthcare, climate, and other mission-critical areas.

Workstream 4: Build talent, capacity, and cross-sector pathways Create an integrated framework for industry exchanges, fellowships, and funding criteria – ensuring sustained pathways for interdisciplinary expertise to shape AI.

Workstream 5: Establish global knowledge infrastructure Create a distributed, open platform for knowledge exchange, field coordination, training, and dissemination – embedding interpretive AI into global research and innovation ecosystems.

Categorisation framework

Domain impact categories	Technical innovation
	Economic value
	Societal benefit
	Global strategic relevance
Timeframe	Short-term (1-2 years)
	Medium-term (2-5 years)
	Long-term (5+ years)
Implementation level	Research activities
	Policy mechanisms
	Industry integration
	Educational transformation



3.2.1 Workstream logic table

#	Focus	Primary audience	Action	Justification	Implementation mechanism	Indicative investment
W1	Interpretive evaluation & representation	Global funders, AI institutes, academic consortia	Coordinate development of assessment frameworks, representation methods, and integration tools	AI systems lack robust mechanisms for representing meaning, context, and cultural nuance	Projects pipeline to co-develop interpretive methods and measures with shared evaluation scenarios	£1M in year 1, £3-5M to year 5, depending on site scope and co-funding
W2	Architectural diversity	Research funders, technical universities	Fund 2–3 interdisciplinary labs to explore system design variations using diverse evaluation approaches informed by humanistic reasoning	Current design patterns limit adaptability; diverse evaluation can incentivise architectural innovation	Labs develop exploratory prototypes using contextual evaluation beyond traditional benchmarks (R1)	£750k–£1M per lab
W3	Collaborative intelligence systems	Interdisciplinary programmes, human-computer interaction & sociotechnical labs, industry partners	Develop ensemble methods that enhance human agency and cultural reasoning in complex systems	Existing models reduce interaction to tools or assistants, missing shared agency and social context	Fund pilots exploring interpretive collaboration in high-stakes, multi-agent settings	£1.5–2M for 5–6 teams
W4	Talent, capacity, and cross-sector pathways	Research institutions, industry partners, funding bodies	Create an integrated framework for talent development and boundary-crossing exchange	Humanities lacks development access; AI lacks interpretive expertise	Industry-academia exchange programme, boundary-crossing fellowships, and institutional incentives with coordinated oversight	£4-6M over 5 years for exchanges, fellowships, and coordination
W5	Gobal knowledge infrastructure	Academic consortia, global institutions, non-governmental organisations	Build open infrastructure to support field formation and long-term collaboration	Field-building requires sustainable, non-extractive models for knowledge sharing and interdisciplinary training	Digital platform, seasonal residencies, curriculum initiatives, and publication partnerships	£500K–£750K/ year for platform + programmes

3.2.2 Detailed workstream descriptions

W1 Develop interpretive AI foundations

Objective 1. Theme	Interpretive technologies
Audience	Global funders, AI institutes, academic consortia.
Action	Support community-led creation of interpretive assessment framework
Justification	Current AI systems simplify or ignore cultural nuance, ambiguity, and contextual depth. To enable richer interpretive reasoning, new forms of evaluation, knowledge representation, and model integration are essential.
Mechanism	Support a pipeline of projects to collaboratively develop plural evaluation measures, context-aware representation methods, and reusable model components for interpretive tasks. This initiative will provide shared evaluation scenarios and guide their integration into research practice.
Technical Advance	Concrete advances in AI's ability to represent and reason about meaning, including benchmarks for interpretive tasks, context-driven training methods, and modular interpretive model elements.
Success Indicator	By 2026, four demonstrator projects funded and launched.
Indicative Investment	£1M in year 1, £3-5M to year 5.
Tags	Technical innovation Societal benefit Short-to-medium term Research activities

W2 Expand AI design pathways through interdisciplinary insight

Objective 2. Theme	Alternative architectures
Audience	Research funders, technical universities, interdisciplinary AI research centres.
Action	Support interdisciplinary labs to explore new system configurations – including design, training, data, and evaluation – informed by humanities-based reasoning.
Justification	Current AI development converges around a narrow set of design patterns optimised for scale. Integrating underused reasoning frameworks can help address performance stagnation and expand system capabilities.
Mechanism	Fund two to three cross-disciplinary labs to develop exploratory prototypes. Projects might integrate, for example, narrative structure in model evaluation, speculative methods in human-AI interaction, or ecological reasoning in training strategies. Outputs evaluated using diverse approaches beyond traditional benchmarks.
Technical Advance	New combinations of established methods and interpretive approaches that improve context handling, expand design logic, or open new use cases.
Success Indicator	By 2028, prototypes demonstrate distinctive capabilities using contextual evaluation methods; frameworks influence AI design practices in academic and applied settings.
Indicative Investment	£750k–£1M per lab; global co-funding encouraged.
Tags	Technical innovation Global strategic relevance Medium-to-long term Research activities

W3 Enable human-AI ensemble systems

Objective 3. Theme Human-AI ensembles

Audience Interdisciplinary research programmes, human-computer interaction, and sociotechnical systems labs.

Action Launch a programme to develop ensemble methods for collaborative intelligence, grounded in interpretive reasoning, shared agency, and cultural context.

Justification Existing approaches to human-AI interaction often reduce human input to feedback or supervision. Without frameworks for mutual adaptation and cultural reasoning, AI systems risk diminishing human capabilities and distorting social dynamics.

Mechanism Fund interdisciplinary teams to create pilot systems and conceptual models that enable dynamic, context-sensitive collaboration between humans and AI. Support development of testbeds and real-world case studies across domains such as sustainability, health, and design.

Technical Advance New paradigms for co-intelligence and interpretive collaboration beyond assistant models, enabling more flexible, socially embedded interaction models.

Success Indicator By 2026: Pilot models and frameworks developed; By 2030: Measurable impact on ensemble-based decision systems and uptake in critical domains.

Indicative Investment £1.5–2M to support five to six pilot teams.

Tags
- Technical innovation
- Societal benefit
- Medium term
- Research activities
- Industry integration

W4 Build talent, capacity, and cross-sector pathways

Objective 4. Field-Building Function Interdisciplinary career pathways

Audience Research institutions, industry partners, funding bodies

Action Establish an integrated framework for cross-sector talent development, including fellowships, placements, and training.

Justification Technical innovation in interpretive and context-aware AI depends on sustained interdisciplinary collaboration and pathways for expertise to flow between academia, industry, public sectors.

Mechanism Targeted fellowship calls, partnerships with industry for exchange schemes, regional representation through global partnerships and residencies, and policy engagement with research funders to establish humanities integration in AI research funding and institutional practices.

Technical Advance Sustained flow of interpretive expertise into system development, auditing, and interface design; emergence of hybrid technical-humanistic methods.

Success Indicator By 2030, an international network of 50+ interdisciplinary leaders in place across academia, industry, and policy; recognised integration of humanities requirements in major AI research programmes and strategic initiatives.

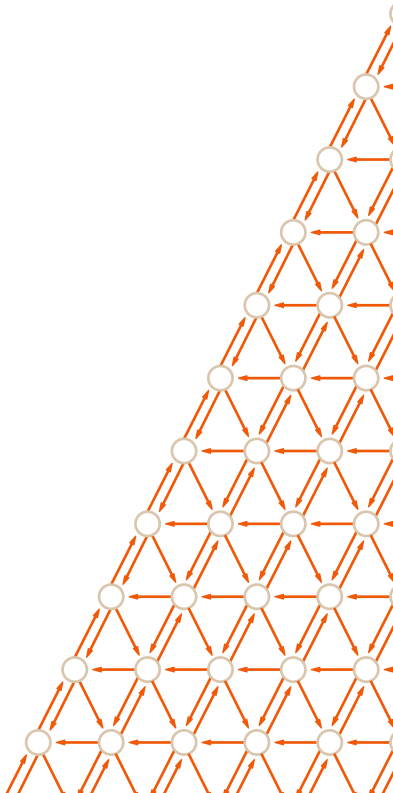
Indicative Investment £4–6M over five years, with global co-funding encouraged from national research councils, philanthropic foundations, and industry partners.

Tags
- Technical innovation
- Economic value
- Long-term
- Industry integration

W5 Establish global knowledge infrastructure

Objective 5. Field-Building Function Epistemic infrastructure and shared culture

Audience	Academic consortia, research institutes, global institutions, funders
Action	Develop an open knowledge platform with supporting infrastructure: seasonal residencies, collaborative resource development, interdisciplinary curricula, and dedicated academic publication outlets.
Justification	Field formation requires sustained, non-extractive, community-led frameworks for knowledge creation and exchange, and training of new interdisciplinary talent.
Mechanism	Digital platform with open resources; residencies and summer intensives hosted across global nodes; collaborative curriculum initiatives; support for journals or special issues focused on humanities-AI integration.
Technical Advance	Shared epistemic tools, curricula, and publication venues enabling durable field-scale knowledge infrastructure.
Success Indicator	By 2026, live platform and convening programme supporting co-creation of resources and community exchange across multiple global regions.
Indicative Investment	£500K–£750K per year for platform + programmes.
Tags	Societal benefit Global strategic relevance Short-to-long term Educational transformation Policy mechanisms



3.3 Implementation mechanisms

Achieving the vision outlined in the implementation pathway requires specific mechanisms that bridge disciplinary boundaries while providing practical structures for collaboration. We propose five complementary mechanisms that enable both immediate impact and sustained development.

3.3.1 Field-building scaffolds

Our methodology shows that enabling structures can bridge individual projects and wider institutional transformation. Formats such as international workshops, contributory authorship, and ongoing community calls have enabled sustained collaboration while remaining responsive to emerging opportunities. Scaling this approach involves distributed convening, collaborative knowledge production, and engagement across research, policy, industry, and public sectors. Future platforms should provide shared infrastructure for long-term inquiry – supporting interdisciplinary work without fixing a single disciplinary frame. This model offers a replicable framework for catalysing similar initiatives on other cross-cutting challenges.

3.3.2 Cross-sector exchange programmes

The gap between humanities expertise and technical development creates a critical barrier to integration. We will establish concrete exchange mechanisms through:

Sabbatical exchanges enabling humanities scholars to embed within technical environments and technologists to engage with humanities contexts.

Industry-academia matchmaking service that identifies complementary expertise and facilitates collaboration.

Joint appointments that institutionalise cross-sector roles and enable sustained engagement.

Co-creation and translation workshops specifically designed to bridge disciplinary vocabularies.

Industrial integration sandpits that enable industry and academia collaboration to trial and de-risk the industrial implementation of innovations developed within Doing AI Differently.

3.3.3 Boundary-spanning roles

Interdisciplinary co-creation requires researchers who can serve as bridges between disciplines, sectors, and methodological approaches. Our experience demonstrates the vital role of facilitators, translators, and brokers who enable productive dialogue across different epistemic cultures.

Key boundary-spanning functions include:

Creative agents who span fields and cultures, acting as cultural and methodological connectors.

"Hinge function" researchers who develop expertise across multiple domains while maintaining credibility in each.

Research translators who bridge academic research with industrial and policy adoption.

These roles require legitimisation and sustainable funding within existing academic and industry structures. Current funding mechanisms often struggle to support positions that don't fit traditional disciplinary categories, despite their essential contribution to breakthrough interdisciplinary research.

3.3.4 Diverse participation pathways

Participation must extend beyond established researchers to build sustainable capacity. Through tiered engagement opportunities, we will create multiple entry points:

Early career fellowships supporting researchers as they establish boundary-crossing expertise.

Research residencies providing intensive immersion opportunities for established scholars.

Artistic residencies inviting creative practitioners to engage with and challenge technical development through embodied and affective approaches.

Summer intensives offering structured training in interdisciplinary methods.

Micro-grants enabling exploratory projects and preliminary collaborations.

Participation pathways will actively support researchers from underrepresented communities and regions, recognising that interpretive AI requires global perspectives to avoid perpetuating narrow cultural assumptions.

3.3.5 Policy and funding alignment

To embed humanities perspectives in AI research at scale:

RRI-inspired funding criteria: Adoption of Responsible Research and Innovation (RRI) style mechanisms in funding calls, requiring humanistic participation in relevant AI projects.

Policy integration models: Collaboration with national and international funders to promote interpretive AI within broader science and technology strategies.

3.4 Barriers and risks

Successful implementation requires acknowledging and addressing key challenges. The table below identifies critical barriers and risks along with specific mitigation strategies.

Challenge	Description	Mitigation strategy
Humanities capacity erosion	Declining funding and institutional support for humanities research limits available expertise and resources.	• Co-funded faculty positions with technical partners • Integration of humanities requirements in AI funding calls • Visible demonstration projects showing value-add
Cross-disciplinary communication	Differing vocabularies, methods, and reward structures between humanities and technical fields create collaboration barriers.	• Translation workshops with shared vocabulary development • Cross-disciplinary mentorship programmes • Recognition frameworks for interdisciplinary contributions
Industry integration	Commercial pressures for rapid development may limit adoption of approaches perceived as adding complexity or cost.	• Demonstrator projects showing improved outcomes • Early engagement with industry partners in design phase • Focus on areas where current approaches clearly fall short
Technical feasibility	Integrating interpretive frameworks into computational systems presents significant technical challenges.	• Incremental technical approaches with defined milestones • Collaborative teams with both humanistic and technical expertise • Realistic timeline expectations for complex integration tasks
Political climate	Shifting political priorities and cultural debates may impact support for humanities-informed approaches to technology.	• Framing that appeals across ideological perspectives • Emphasis on practical outcomes and competitive advantage • Diversified funding sources across sectors and regions
Creative sector displacement	Generative AI has already impacted artists and cultural workers through uncompensated appropriation and economic precarity.	• Establish clear frameworks for attribution, licensing, compensation, and consent • Support "human-made" provenance tools • Mandate disclosure of artistic data usage • Invest in artist-inclusive AI governance and co-design roles
Commercialisation barriers	High cost of model development (£ billions) creates significant barriers to entry for new approaches.	• Focus on influencing existing development pipelines • Targeting specific applications where value is demonstrable • Building partnerships with established industry players
Risk of misuse	Sophisticated interpretive AI could be exploited for persuasive manipulation, sophisticated disinformation, cultural exploitation, or invasive monitoring.	• Ethics-by-design framework integrated into model development • Adversarial testing protocols • Independent oversight mechanisms • Parallel detection technologies to identify misuse.
Implementation scale	Ensuring approaches scale beyond academic settings to impact industry practices at meaningful levels.	• Design for scalability from early stages • Documentation of methodologies for wider adoption • "Train-the-trainer" approaches to amplify impact

This framework acknowledges significant challenges while providing concrete paths forward. By anticipating these barriers and risks early in implementation, we can design more resilient approaches and set realistic expectations for progress.

3.5 Funding model

Delivering all five workstreams will require sustained support over multiple years. Funding will need to support collaborative labs, new infrastructure, cross-sector mobility, and an active international community of practice.

We estimate that funding in the order of £10M to £20M over five years can be expected to take us toward the goals described here. These figures are indicative and subject to refinement based on delivery model, co-funding arrangements, and partner priorities.

To support different routes to implementation, we outline two illustrative delivery models below:

Rapid Delivery: Concentrates effort in a small number of large projects, based in well-established institutions. Prioritises early demonstrators, fast delivery of benchmarks, and centralised coordination.

Strong Roots: Distributes investment across a wider set of partners, with a focus on regional participation, community-building, and support for new or underrepresented entrants. Enables slower but broader field growth.

These models are not mutually exclusive and may be combined in phased or hybrid approaches.

This estimate is consistent with investments used to establish comparable interdisciplinary fields – for example, the £10M founding investment that launched Data-Centric Engineering (DCE) at The Alan Turing Institute. Similarly, the EU allocated over €460M to Responsible Research and Innovation (RRI) initiatives during Horizon 2020.

3.6 Success metrics and timeline

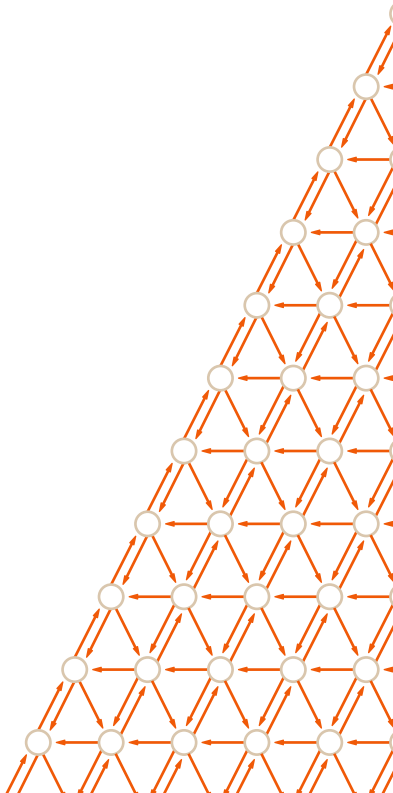
Success measures

Area of transformation	Technical metrics	Economic indicators	Policy alignment
Interpretive foundations	• 3-5 new evaluation methods adopted by research community • Integration into 2+ model evaluation frameworks	• Applied in 2+ commercial or public AI systems • £5-8M in collaborative research and development partnerships	• Referenced in global AI policy frameworks • Integrated into technical and ethical standards
Architectural innovation	• 2-3 prototype systems outperforming traditional models on interpretive tasks • 5+ peer-reviewed publications in top computer science venues	• 1-2 spin-outs or industry partnerships • Adoption in open-source or enterprise systems	• Support for technical plurality in international standards bodies • Inclusion in multinational research funding priorities
Collaborative systems	• Pilot ensemble methods tested in 3+ domains • Demonstrable improvements in task performance	• Productivity enhancements in test environments • Integration into enterprise and public platforms	• Alignment with human-centred AI policies • Referenced in responsible AI initiatives
Cross-Sector capacity	• 20+ skilled professionals moving between sectors • New hybrid methods documented and shared	• New job categories and roles established • Market growth in intermediary services	• Embedded in research and education frameworks • Integrated into global workforce development strategies

Short, medium, and long-term outcomes

Note: Projected impacts depend on securing investment in the indicative range of £10–20M over five years, with scope for phased or regionally varied delivery depending on the model adopted.

Workstream	Short-term outputs (1 year)	Medium-term outcomes (2–3 years)	Long-term impacts (5+ years)
W1 Interpretive frameworks & measures initiative	4 demonstrator projects launched; benchmarks and contextual methods in early development	Plural evaluation measures and model components in research use	Evidence that AI systems can engage with cultural contexts without undermining community self-determination or displacing human interpretive practices
W2 Expand AI design pathways	First labs convened; initial design directions established	Prototypes demonstrate value of integrated reasoning approaches	Humanities-informed design logics shape new technical pathways in AI, leading to greater system adaptability and contextual robustness
W3 Ensemble methods programme	Experimental systems and protocols launched in pilot domains	Toolkits and methods libraries in use; performance gains documented	Healthcare teams use AI that amplifies clinical judgment, improving care quality
W4 Talent and cross-sector pathways	Programme launched; Funders engaged.	First cohort completed; case studies published	Humanities and arts expertise is embedded upstream in AI development processes across public and private sectors
W5 Global knowledge infrastructure	Collaborative forum established, with international participation and plans to deepen regional anchoring	First residencies and summer intensives hosted; active community sharing tools, curricula, and research outputs	Interpretive AI becomes a globally networked field with accessible shared tools, knowledge platforms, and open participation



3.7 Why we believe this can succeed

This initiative builds on proven methodology that has already achieved significant impact within 18 months, supported by demonstrated institutional capacity and leadership experience:

Demonstrated success:

Systematic community mobilisation: Design-led approach successfully engaged 150+ researchers across six continents – co-authoring papers, organising workshops, and building shared foundations for an emerging field.

Rapid institutional adoption: Generated £1M+ in funding and policy integration, indicating genuine alignment with strategic priorities.

Community-endorsed vision: The agenda has been co-developed with and validated by 70+ experts, through a rigorous consultation process culminating in the March 2025 workshop.

Established infrastructure and leadership:

Established convening power: The Alan Turing Institute, University of Edinburgh, Arts & Humanities Research Council, and their partners have jointly convened over 40 institutions across the humanities and AI – providing coordination capacity and transdisciplinary leadership at scale.

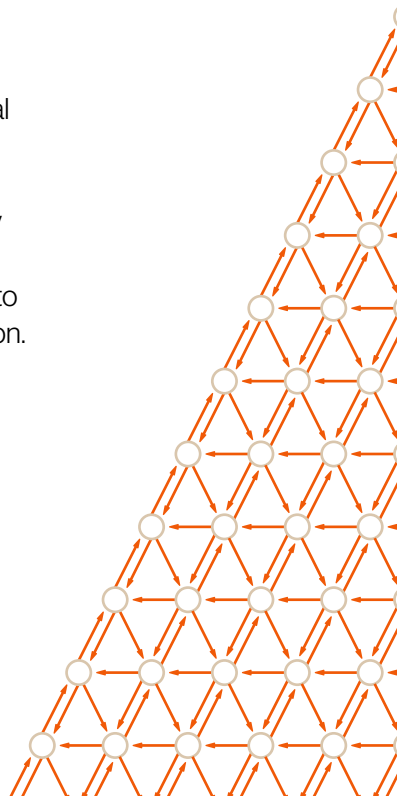
Proven policy influence: We have already secured policy adoption through a programme theme at Arts & Humanities Research Council, resulting in a landmark £1M funding call with Social Sciences and Humanities Research Council for UK-Canada collaborative projects in this space.

Track record in innovation: The team leading this initiative has successfully established and/or led major new research domains, including Data-Centric Engineering at the Turing Institute, and AI Arts through The New Real and FutureEverything.

Global collaborative network: Active partnerships with over 40 institutions and organisations across six continents are enabling distributed experimentation, regional anchoring, and cross-sector collaboration.

Strategic alignment: This initiative builds on growing global consensus that AI must evolve to reflect cultural and contextual awareness – as seen in OECD, UNESCO, and national AI strategies worldwide. Doing AI Differently offers a technically credible and globally relevant starting point for rethinking AI foundations in ways that honour cultural complexity and human interpretive agency.

These foundations demonstrate that the opportunity is real and momentum is already building. Doing AI Differently is not owned by any single institution, but is a platform for collective leadership – open to all researchers, practitioners, and organisations ready to shape the role of humanities in AI's future a critical moment in the technology's evolution.



4. Collaborative next steps

The challenges and opportunities identified in this white paper demand coordinated action across research, policy, and industry. This final section outlines immediate steps that different stakeholders can take to advance this initiative and create momentum for change.

- Join the community of research partners in 2025 (research leaders).
- Participate in the International Science Partnerships Fund (ISPF) Sandpit and Funding Call launching Summer 2025 (research community).
- Participate in the first industry-academia exchange programme (industry).
- Contribute domain expertise for case studies and testbeds (industry, institutions).
- Co-develop funding and capacity-building pathways to support interdisciplinary talent and research (funders, institutions).
- Incorporate interpretive AI requirements into funding calls (funders).
- Contribute to the global knowledge platform through research, case studies, and applied insights (all).

Signatories to the Doing AI Differently White paper

Hannah Andrews (Director of Digital Innovation in the Arts, British Council)
Paul Batterham (Chief Innovation Officer, Kainos)
Christina Boswell (Vice President for Public Policy, British Academy & Vice-Principal Research and Enterprise, University of Edinburgh)
Nick de Pennington (Founder CEO, Ufonia)
James Flint (Loughborough University/Institution of Engineering and Technology)
Marc Funnell (Director for Digital Engineering, High Value Manufacturing Catapult)
Tarleton Gillespie (Senior Principal Researcher, Microsoft Research)
Maya Indira Ganesh (Associate Director, Leverhulme Centre for the Future of Intelligence)
Ben Harris (Partner, Newton)
Owen Hopkin (Director of New Technologies & Innovation, Arts Council England)
Aisha Iqbal (AI Policy and Governance, Meta)
Roger F. Malina (President, Association Leonardo)
Arturo Muente-Kunigami (Digital Transformation, Data Policies and AI, Inter-American Development Bank)
Pierre-Yves Oudeyer (Research Director, Inria)
Jan Przydatek (Director of Technologies, Lloyd's Register Foundation)
Sally Radwan (Member of the Board of Trustees, The Alan Turing Institute)
Raul Santos Rodriguez (Director, Bristol Digital Futures Institute)
Allan Sudlow (Director of Partnerships and Engagement, Arts and Humanities Research Council)
Krystyn J. Van Vliet (Vice President for Innovation & External Engagement Strategy, Cornell University)
Steph Wright (Head, Scottish AI Alliance)
Lili Jia (Research Professor, Cambridge Centre for Human Inspired AI)

The Doing AI Differently initiative is led by The Alan Turing Institute, University of Edinburgh and the UK's Arts & Humanities Research Council (AHRC-UKRI), in collaboration with international partners.

Publisher: The Alan Turing Institute, London, UK.
Publication Date: 31 July 2025.
DOI: 10.5281/zenodo.16421296
Licence: CC BY 4.0.

Cite as: Hemment, D., Kommers, C., et al. (2025). Doing AI Differently: Rethinking the Foundations of AI via the Humanities. White Paper. London: The Alan Turing Institute.

Funded by: Arts & Humanities Research Council and Lloyd's Register Foundation.

The white paper and accompanying policy note and methodology report are available online: www.turing.ac.uk/news/publications/doing-ai-differently

Doing AI Differently website: www.doingaidifferently.org

Cover image: Yutong Liu & Kingston School of Art / Better Images of AI / Talking to AI 2.0 / CC-BY 4.0.

Design and typography: Joshua Smythe, Studio Ampersandwich

Brand identity: Steven Scott, twofifths design

About The Alan Turing Institute
The Alan Turing Institute is the UK's national institute for data science and artificial intelligence. The Institute is named in honour of Alan Turing, whose pioneering work in theoretical and applied mathematics, engineering and computing is considered to have laid the foundations for modern-day data science and artificial intelligence. The Institute's purpose is to make great leaps in data science and AI research to change the world for the better. Its goals are to advance world-class research and apply it to national and global challenges, build skills for the future by contributing to training people across sectors and career stages, and drive an informed public conversation by providing balanced and evidence-based views on data science and AI.

The Alan Turing Institute
THE UNIVERSITY of EDINBURGH
UKRI Arts and Humanities Research Council
LR Foundation