HACID - Deliverable

Project catalogue

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101070588. UK Research and Innovation (UKRI) funds the Nesta and Met Office contributions to the HACID project.

Deliverable number:	D8.6
Due date:	31.08.2025
Nature[1]:	R
Dissemination Level[2]:	PU
Work Package:	WP8
Lead Beneficiary:	Nesta
Contributing Beneficiaries:	ISTC - CNR

[1] The following codes are admitted:
- R: Document, report (excluding the periodic and final reports)
- DEM: Demonstrator, pilot, prototype, plan designs
- DEC: Websites, patents filing, press & media actions, videos, etc.
- DATA: Data sets, microdata, etc.
- DMP: Data management plan
- ETHICS: Deliverables related to ethics issues.
- SECURITY: Deliverables related to security issues
- OTHER: Software, technical diagram, algorithms, models, etc.

[2] The following codes are admitted:
- PU – Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page)
- SEN – Sensitive, limited under the conditions of the Grant Agreement
- Classified R-UE/EU-R – EU RESTRICTED under the Commission Decision No2015/444
- Classified C-UE/EU-C – EU CONFIDENTIAL under the Commission Decision No2015/444
- Classified S-UE/EU-S – EU SECRET under the Commission Decision No2015/444

Document History

Version	Date	Description	Author	Partner
0.1	19.08.2025	Creation	Aleks Berditchevskaia	Nesta
0.2	26.08.2025	First draft	Aleks Berditchevskaia, Alexandra Albert, Christopher Edgar, Ewa Dominiak, Rita Marques	Nesta
0.3	27.08.2025	Partner Review	Vito Trianni	ISTC - CNR
0.4	28.08.2025-29.08.2025	Revisions	Aleks Berditchevskaia, Alexandra Albert	Nesta
0.5	29.08.2025	Final Review	Peter Baeck	Nesta

Table of contents

1. Introduction to the project catalogue

This project catalogue provides key communication and design assets that highlight the value of combining AI with collective intelligence to support decision-making in high-stakes domains. The catalogue details insights and recommendations gleaned from three policy roundtables and presents blueprints for two alternative applications of the HACID approach in the energy and water sectors.

The core of our work demonstrates that a hybrid approach - one that integrates the power of technology to analyse vast datasets with the contextual and tacit knowledge of human experts - can lead to more effective, transparent, and trusted outcomes. We show how this methodology can not only enhance the performance of decision support systems but also address critical issues of accountability, safety, and public trust.

This document is designed to inspire and inform policymakers, regulators, and practitioners. By showcasing concrete use cases and practical, actionable recommendations, we aim to demonstrate how a HACID-inspired approach can be exploited in diverse domains facing complex, uncertain challenges. It provides a foundational resource for organisations looking to build safer, more inclusive, and publicly accountable hybrid intelligence systems.

Policy Roundtables: Navigating high-stakes domains

The catalogue summarises insights from three expert roundtables focused on the practical challenges of embedding responsible AI. These discussions explored:

1. **Practitioner Perspectives on Participatory AI:** This roundtable focused on the practicalities and challenges of involving a wider range of stakeholders in the development and governance of AI. It explores why meaningful participation is essential for improving AI performance, increasing usability, and building trust.
2. **AI Regulation for Clinical Decision Support:** In the high-risk field of medical diagnostics, this roundtable addressed the operational and institutional barriers to AI adoption, the role of regulation in bridging the gap between principles and practice, and how to build trust through inclusive design and oversight.
3. **AI Governance and Trust in Climate Services:** This roundtable explored the unique challenges faced by national organisations, such as the Met Office, in deploying AI for public good. Key themes included verification, public trust, and the need for new institutional structures to ensure that AI-enabled services align with values like fairness and transparency.

Alternative Use Cases: Extending the HACID Approach

Building on these insights, the catalogue presents two alternative applications of the HACID methodology in domains where decision-making is complex and has significant consequences.

1. **The Energy Sector:** We explore how a hybrid intelligence tool can support community-led energy initiatives. This domain is particularly suitable for the HACID

approach because it requires navigating immense uncertainty, integrating diverse data sources (from weather patterns to community feedback), and democratising decision-making to ensure equitable access and long-term sustainability.

2. **Water Regulation:** This use case demonstrates how a hybrid system can assist regulators like Ofwat in the UK in making more effective, transparent, and inclusive decisions. The water management sector is characterised by fragmented data, distributed expertise, and the need to proactively identify and manage emerging risks, making it an ideal candidate for an approach that combines AI's analytical power with human contextual knowledge.

The findings from these roundtables and use cases collectively demonstrate that a responsible and trustworthy approach to AI is not an afterthought, but a core component of its successful, long-term adoption. By providing a framework for incorporating diverse perspectives, clarifying accountability, and focusing on real-world problems, this catalogue offers a blueprint for building the next generation of hybrid intelligence systems.

2. Policy Roundtables: Navigating high-stakes domains

To allow technologies like HACID to thrive and scale, it is important to establish an enabling policy and funding ecosystem. We hosted 3 roundtables targeted towards the policy audiences and other relevant stakeholders for each of the two existing HACID use cases and AI innovation for high stakes domains (particularly public sector uses). Through these roundtables we identified how technologies like HACID could fit into existing programmes and address current and future policy objectives, as well as stress-testing emerging findings and recommendations about how to implement, support and scale HACID.

2.1 Policy Roundtable - Practitioner perspectives on Participatory AI

Why Participatory AI Matters for Public Trust and Legitimacy

Participatory AI in its broadest sense, refers to the involvement of a wider range of stakeholders than just technology developers in the creation (and governance) of an AI model, tool or application.[3] The benefits of using participatory approaches range from performance related improvements in accuracy, through to justice-based considerations that aim to challenge existing power structures in the development of technology.[4] Existing case studies using participatory AI have reported the following benefits:

---

[3] https://www.nesta.org.uk/project/participatory-ai/. Accessed 20/08/2025.
[4] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23), https://doi.org/10.1145/3617694.3623261

- It helps to improve model performance;
- It increases the usability, appropriateness and uptake of a tool for a given problem;
- It helps to align the tool with diverse values, needs and preferences;
- It can assist with anticipating and mitigating broader impacts, risks or harms; and
- It increases trust in AI by different stakeholders and trustworthiness of tools.

The roundtable, held in April 2024 at Nesta's offices in London, in collaboration with the Ada Lovelace Institute,[5] brought together a range of stakeholders from industry and technology companies, and from local governments. It aimed to explore existing methods within Participatory AI (PAI) at a high level, and to discuss the practical constraints and realities of institutionalising PAI in industry. For example, ethnographic studies of technology teams have highlighted the tensions between commercial incentives for adopting fairness-aware, responsible design approaches[6] and the gap between identifying specific issues and knowing how to address them.[7]

The roundtable also aimed to identify and prioritise new opportunities for participation during AI development and for participants to share experiences between different organisations related to PAI since this is a broadly nascent practice. In the last year there has been a wave of enthusiasm for participation, driven by major technology companies including Meta[8] and OpenAI.[9] Ostensibly this is to address ethical concerns relating to generative AI technology and LLMs. This roundtable was the first conversation, hopefully of many, to assess whether the claims made for participation are true and how to implement methodologies in a way where these aims can be realised.

The discussions and key takeaways are summarised below. We conclude with recommendations relevant to people working in industry or technology companies, ideally people who are engineers or machine learning scientists. They are also for anyone working in tech teams in public sector organisations at both the local and national levels.

Section 1 World Cafe Activity

Through a World Cafe activity, participants were asked to reflect on how they might apply participatory methods to their work, whether that is a data or an AI project or something else. The use of a world cafe approach, where participants contributed post-it notes to stimulate discussions in the room, meant that we do not have direct quotes from participants in this section of the catalogue.

---

[5] Ada Lovelace Institute is an independent research institute in the UK, funded by the Nuffield Foundation. Since the institute was founded they have maintained an active programme of work on Public Participation and Research to ensure "voices of people affected by data and AI contribute to building and shaping evidence, research, policy and practices."
https://www.adalovelaceinstitute.org/our-work/research-domains/public-participation-research/

[6] Lara Groves, Aidan Peppin, Andrew Strait, and Jenny Brennan. 2023. Going public: the role of public participation approaches in commercial AI labs. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 1162–1173

[7] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé, Miro Dudik, and Hanna Wallach. 2019. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). https://doi.org/10.1145/3290605.3300830

[8] https://deliberation.stanford.edu/meta-community-forum-generative-ai-results

[9] https://openai.com/index/democratic-inputs-to-ai/

Image 1: participant post-its from the World Cafe activity

Discussion summary
We structured the World Cafe discussions around 3 key topics:

1. What are the relevant organisational roles related to Participatory AI development within an organisation? What priorities do they have in relation to PAI?
2. What are the incentive structures for developers and participants in relation to PAI?
3. What barriers and enablers exist across industry to the institutionalisation of PAI?

We also offered a wild card table to capture any additional topics raised by participants but we found that the other topics provided enough scope to capture the priority themes that attendees wanted to discuss.

Key takeaways

Need to clarify responsibility and accountability lines for Participatory AI: Ownership of participation in large tech companies (with employee numbers in the thousands) is difficult, as participatory activities will likely be the assigned responsibility of a given team (e.g. UX research). These responsible teams design and implement PAI activities and then synthesise recommendations to be forwarded to management, developers, and data scientists or machine learning engineers. Thus, the challenge is ensuring that appropriate actions are taken rather than getting lost between those doing the research and other parts of the company, with clear accountability.

PAI champions to communicate and advocate on Participatory AI: There is a challenge related to the communication of findings, and advocacy for the use of PAI at all. To address this, it is crucial to have specific individuals responsible for oversight and communication around PAI - 'PAI champions' within an organisation. PAI champions are also appealing to the local government perspective, with the emphasis on communication and advocacy in the form of knowledge sharing between councils - as they seem to work in a very siloed fashion.

Incentive structures for developers *and* participants: The aims of participation are at odds with the way priorities are set in technology companies which are made on the basis of company priorities and strategies. There is a lot of competition between priorities, strategies, and areas of focus in the fast-paced industry of technology development. This is also true in the public sector, where local authorities have to compete to demonstrate their data/AI capability. Furthermore, the liability for errors or harms is unclear at different stages of the pipeline. **Reframing participation in terms of profitability** could help incentivise companies. However, since developers and designers want to build ethical and good tools, it is crucial to get them on board with PAI, and to incentivise model developments with an emphasis on functionality and safety. In terms of incentives for products/tools, emphasis should be placed on better user experience and marketing, to also enhance the experience of participants.

Participation is frequently perceived predominantly through the lens of UX design, by people in technology development, where "participation" is frequently done in the form of A/B testing. This involves getting large amounts of micro-feedback to make incremental design improvements to tools. However, some of those working in technology development struggle to interpret richer qualitative feedback that doesn't directly translate to a technological fix. One of the key challenges to address is in streamlining the PAI process, so there is not only careful design of the participatory activities, but sufficient thought is invested in guidance about how to implement these activities and act on the outcomes from the outset.

Section 2: Hack the PAI framework activity

In the next activity participants were invited to hack the PAI framework. In 2021 we published a briefing paper where we mapped out the work on PAI - to create a draft framework that captures the different types of interventions available during the model design and development process, taking examples from the literature and best practice guides in technology design. Participants were shown an early version of the PAI framework and asked to reflect on the value of having such a framework to guide the participatory design of technology, so as to get a sense for which are the most and least promising participatory interventions, in practice.

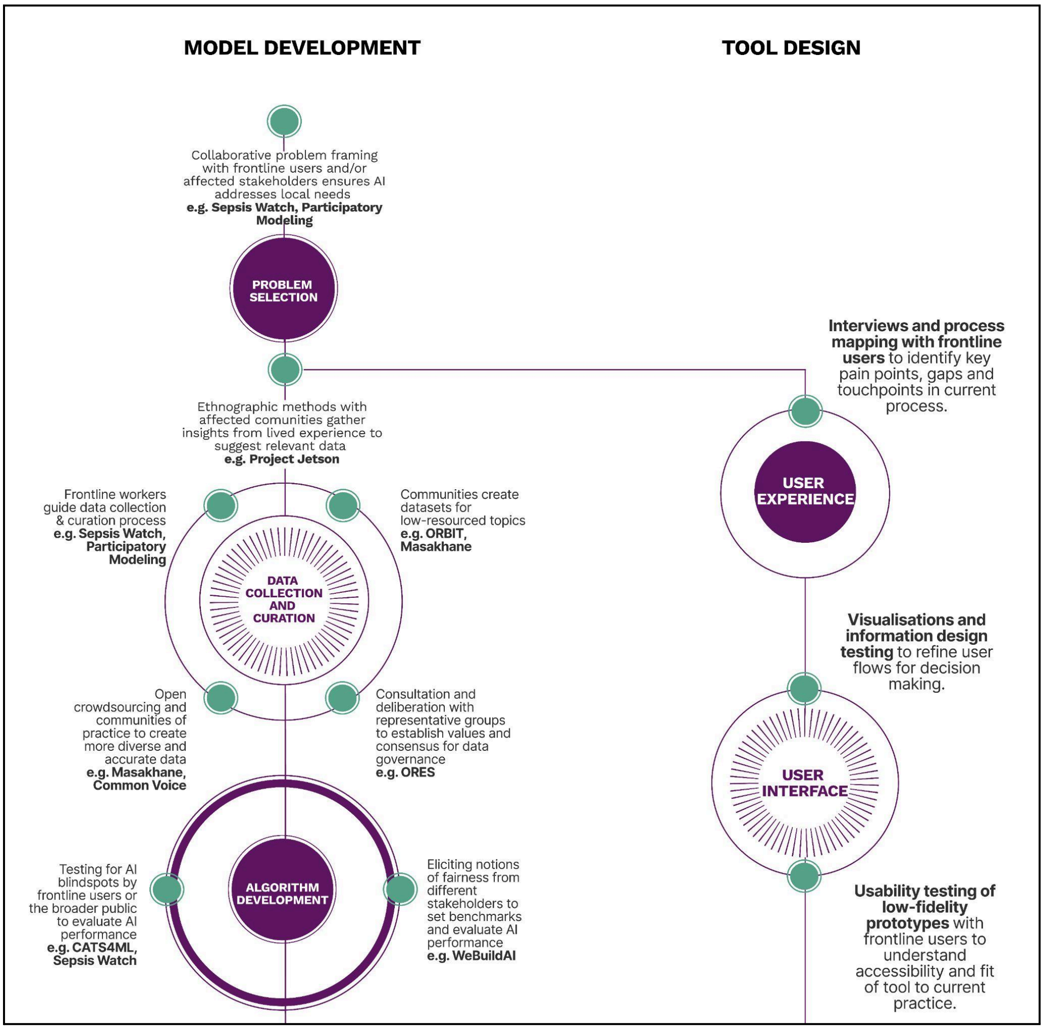


Figure 1: The PAI Framework: an overview of different approaches throughout tool development

On the left hand side of Figure 1 is a simplified linear version of the data collection and AI development pipeline. This would typically be referred to as the backend of the tool, where the focus is on alignment and trust in the model development. On the right hand side of Figure 1 is the tool frontend where the tool interface and user experience takes place. The focus here is on usability. Interventions include validation, verification steps that can be thought of as more comprehensive auditing of models or curation of datasets. Other steps are about identifying the values that are important for stakeholders or identifying the risks they're most worried about. Outputs from these types of activities can be implemented into the design at various stages. We set out to test whether a participatory approach could result in both improvement of the accuracy and appropriateness of an AI-driven tool, as well as explainability and trust.

Participants were asked to use participatory intervention cards as examples from our framework, or to add different examples, and to select an intervention and discuss in small groups where it should be placed on the matric in Figure 2. The matrix has an axis for feasibility: hard to easy to implement; and an axis for impact - how impactful would this intervention be in the development process.

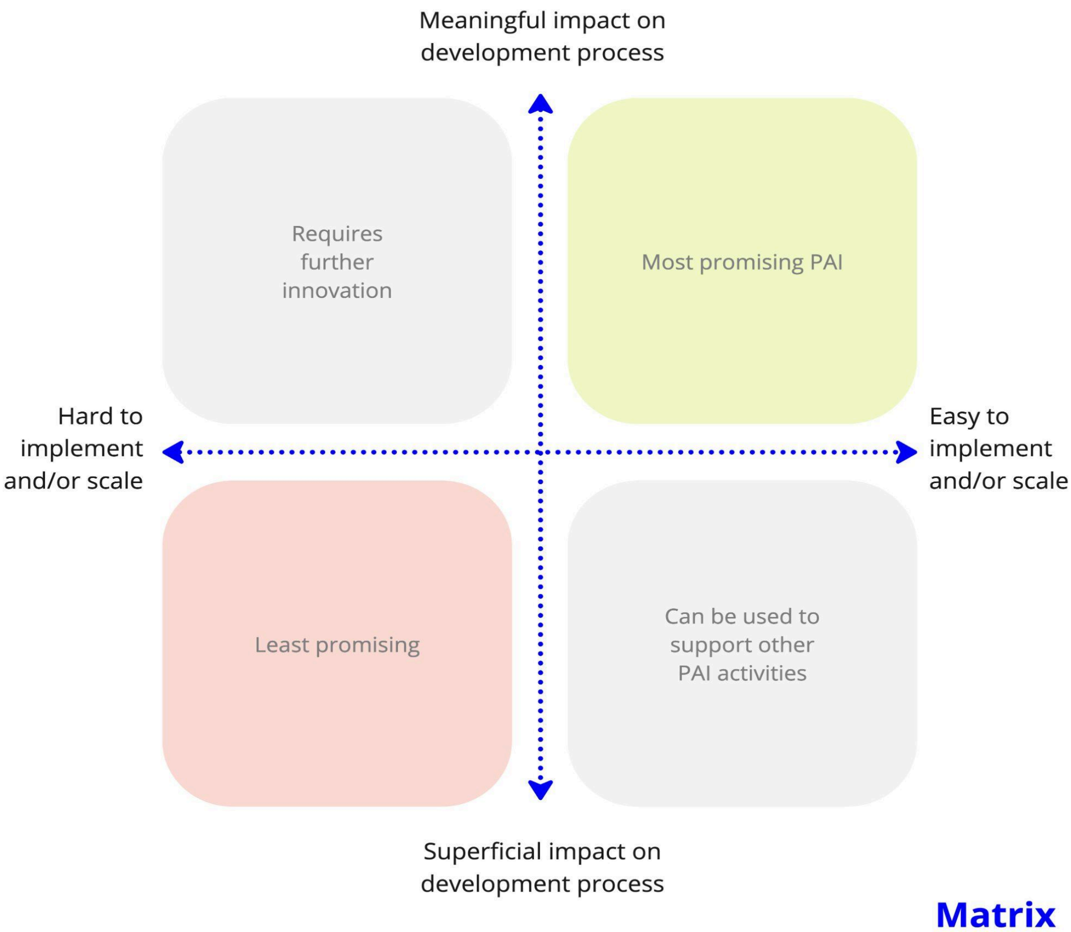


Figure 2: Participatory AI Intervention Assessment Matrix used in the Hack the Framework Activity.

Key takeaways

- Interventions earlier in the pipeline were perceived as more impactful.
- For collaborative problem definition to be successful, it needs both people with technical expertise and problem holders to come together. These groups hold equally important expertise - such as what kinds of problems can be solved with Machine Learning / AI, and what kind of problems need solving in a given domain.
- Participants struggled to find an intervention for the “Least promising” quadrant in Figure 2. This could be due to a lack of AI sceptics in the room.
- Interventions on UI/user-centred design (the usability focus of the framework) are business as usual in both tech companies and government.
- Many of the activities were perceived as having states of being done well (difficult but impactful), and not being done well (fast/easy but not so impactful). Meta’s recent deliberative exercises were held up as a typical example of the ‘easy, quick, and not so impactful’ version of these activities, whereas the Wikimedia foundation’s work on building a relationship with communities to sustain their contributions and participation over time, was perceived as being impactful over time.
- Community-based or deliberative approaches were seen as ‘gold standard’ but often beset by a number of quite fundamental challenges, such as the need for significant amounts of funding, that could mean they lose their poll positioning.

Conclusion and Emerging Recommendations

These recommendations are for people working in industry or technology companies, ideally people who are engineers or machine learning scientists. They are also for anyone working in technical teams in public sector organisations at both the local and national levels.

- **Develop new ways to embed early interventions:** The roundtable found that involving stakeholders earlier in the development pipeline, particularly during the problem definition and data collection stages, is more impactful than later interventions. This will require creative and novel public engagement methods that are aligned with the realities of current technology development, where technology pipelines increasingly rely on foundation models rather than novel data collection.
- **Establish "PAI Champions":** Technical teams should designate specific individuals to be "PAI Champions." This role is crucial for advocating for participatory methods, communicating findings from engagement activities, and ensuring accountability for how participatory AI activities are implemented and acted upon.
- **Develop guidance for senior leaders in local government to demonstrate the value of responsible AI, particularly Participatory AI.** It should emphasise the role of participatory approaches in building public trust and increasing the usability and uptake of new tools, using specific examples.
- **Mandate and incentivise participation:** Public sector commissioners of AI tools should consider making participation a required part of the development lifecycle for AI systems in high-stakes public sector domains. This could involve mandating that technical teams demonstrate stakeholder involvement at specific project stages, and ensuring this is reflected in procurement processes.
- **Encourage cross-government learning:** Regulators should create forums or channels for public sector technical teams to work through specific use cases or concrete examples to consider PAI activities. Use these exchanges as a catalyst for developing good practice, easy-to-use guides, for running participatory AI activities, with a clear role for technical teams.[10]

2.2 Policy Roundtable - AI Regulation for Clinical Decision Support – From Principles to Implementation

Why AI Regulation in healthcare matters

Artificial intelligence (AI) is rapidly transforming healthcare, from diagnostic support and population health management to drug discovery and personalised treatment. However, alongside its promise comes profound risk. When AI tools are deployed in clinical settings without appropriate safeguards, the consequences can include harm to patients, algorithmic bias, loss of trust, and systemic inequality in access and outcomes. Regulatory frameworks

[10] Partnership for AI's Guidance for Inclusive AI is one initiative that attempts to bring together guidance on participatory approaches across different roles. (https://partnershiponai.org/guidance-for-inclusive-ai/)

play a critical role in ensuring that these technologies are safe, transparent, and used responsibly—especially when influencing clinical decision-making.

Policymakers worldwide are accelerating their response. The EU's AI Act—classifying health applications as "high-risk"—and the UK government's AI Opportunities Action Plan reflect growing consensus that cross-sector strategies must be complemented by domain-specific regulations in safety-critical areas like healthcare.[11,12] Guidance from the UK Medicines and Healthcare products Regulatory Agency (MHRA), the US Food and Drug Administration (FDA), and Health Canada through the Good Machine Learning Practice principles is a step in that direction.[13] Yet gaps remain—not only in regulation, but in coordination, evaluation, infrastructure, and public engagement.

Recent research by the Health Foundation shows that while both the public and NHS professionals are broadly supportive of AI, they also express concerns about transparency, accountability, and AI's impact on the human dimension of care. Over half of surveyed NHS staff said AI risks distancing them from patients, and a significant portion of the public remain unsure about how AI decisions will be explained or challenged.[14] These findings underscore the need for tailored, trustworthy regulatory approaches that address the realities of clinical practice and centre patient and practitioner perspectives.

This policy roundtable was delivered in partnership with the UK's AI4CI Hub,[15] a UKRI National AI Research Hub involving over forty stakeholder partners from across academia, government, charities and industry to generate research capacity and practical applications in AI for collective intelligence research across the UK.

We brought together regulators of medical technology, academics, experts in ethical AI, think tanks and public sector organisations (see Appendix 4.2 for a list of participating organisations) to explore three core topics:

- Operational and Institutional Barriers to Embedding AI in Clinical Decision-Making
- Bridging the Regulation Gap – From Principles to Practice
- Participatory Design and Oversight – Building Trust Through Inclusion

The discussions and key takeaways are summarised below. We conclude with recommendations relevant to regulators and institutions deploying AI technology in healthcare contexts.

Section 1: Operational & Institutional Barriers to Embedding AI in Clinical Decision-Making

The roundtable began with a discussion of the key operational and institutional obstacles facing the safe, consistent adoption of AI in clinical decision support. Participants highlighted

[11] Act, E. A. I. (2024). The EU Artificial Intelligence Act.
[12] AI Opportunities Action Plan, Department for Science, Innovation & Technology, accessed 13/05/2025
[13] Busch, F., Geis, R., Wang, Y. C., Kather, J. N., Khori, N. A., Makowski, M. R., ... & Bressem, K. K. (2025). AI regulation in healthcare around the world: what is the status quo?. *medRxiv*, 2025-01.
[14] AI in health care: what do the public and NHS staff think?, The Health Foundation, accessed 13/05/2025
[15] https://ai4ci.ac.uk/

several interconnected themes: fear and misunderstanding of AI technologies, unclear lines of professional liability, a lack of training and strategy, and deep variability in digital maturity across institutions. Together, these factors are limiting effective implementation and fostering mistrust among practitioners.

Key Takeaways from This Discussion

- AI mistrust stems from both cultural narratives and structural gaps - particularly the absence of clear strategies or organisational frameworks.
- Accountability remains murky, particularly in shared human-AI decision-making contexts. Fear of liability inhibits adoption.
- Training and team redesign are essential; AI adoption is not plug-and-play - it reshapes roles, responsibilities, and workflows.
- Digital inequality across healthcare settings threatens to entrench systemic inequities unless infrastructure is upgraded and standardised.
- Confidence in AI systems is undermined by poor-quality data, lack of transparency, and unclear governance structures.

Main Discussion points

Cultural and Institutional Resistance: A pervasive fear of AI persists across the healthcare system, fuelled by sensationalist media narratives and the absence of clear national strategies. These cultural dynamics shape both public and professional perceptions, often leading to misunderstanding, mistrust, and stalled adoption.

“We're struggling really badly with an understanding and a fear factor of AI... The adoption levels are really low because of a fear factor.”

Responsibility, Accountability, and Legal Ambiguity: Clinicians are unsure who holds liability when AI tools influence decisions, creating fear around adoption. The lack of institutional guidance on shared responsibility leaves professionals exposed and hesitant to engage.

Workforce Capacity and Multidisciplinary Integration: AI adoption requires not just technical expertise, but new roles and reconfigured teams. Skills gaps and limited digital fluency among clinicians remain a major barrier to effective implementation.

“People are using old forms of knowledge to apply to new systems... people are completely misunderstanding the concept of AI in their practice.”

Digital Maturity and Access to high-quality, well-governed data remain a significant barrier: Poor-quality data and inconsistent digital infrastructure undermine confidence in AI tools. Wide disparities in digital maturity across healthcare settings block integration and limit scalability.

Lack of central authority on AI use results in a systemic “loss of control”: AI disrupts traditional clinical hierarchies, introducing uncertainty about who holds decision-making authority. This contributes to a broader sense of systemic “headlessness” and weakens trust.

"We don't have the right bodies or the right lines of communication to give people confidence that there is a locus of control that will fix AI, and that contributes to this sense of headlessness and a lack of confidence."

Section 2: Bridging the Regulation Gap – From Principles to Practice

The second discussion focused on how to make AI regulation in healthcare more actionable, addressing the often-cited tension between innovation and safety. Participants agreed that while this tension is real, it is not insurmountable—and may be overstated if regulation is well-designed, anticipatory, and aligned with actual needs. Central to the discussion were regulatory sandboxes (reflecting the extensive input from MHRA), the importance of guidance over rigid enforcement, and the need to build system-level coherence across government, developers, and healthcare delivery bodies.

Key Takeaways from This Discussion

- Innovation and safety can be mutually reinforcing, but only with proactive, transparent regulation and sector-wide alignment.
- Regulatory sandboxes are useful but insufficient alone - they must be part of a broader ecosystem that includes scalable support, stakeholder engagement, and anticipatory regulation.
- Speed vs. safety is not just technical - it's cultural. Regulatory tools must bridge the values and timelines of both tech and healthcare sectors.
- Education and awareness of regulatory context are critical - developers must understand the healthcare environment, and regulators must coordinate to reduce fragmentation.
- Good guidance is often more impactful than rigid rules, especially in rapidly evolving domains like AI.

Main Discussion points

AI as Regulated Medical Software: AI is currently governed under existing medical device regulations, with evolving tools like pre-change control plans addressing ongoing updates, risk, and model transparency.

Innovation and Safety as Reinforcing Goals: When regulation provides clear, anticipatory guidance, innovation and patient safety can be mutually supportive rather than in conflict.

"It's everybody having [an] understanding of what's possible, what's not possible. What is the need that you're actually trying to address? …the more knowledge and information is spread… the more certainty and competency we create, and then the innovation will happen."

Regulatory Sandboxes as Partial Solutions: Sandboxes offer valuable controlled environments for testing AI tools but remain resource-intensive and insufficient without broader infrastructure and incentives.

“We want change. But we want [to do it] in a controlled way. So this is why we have the sandbox. So we can test what systems work and what doesn't work and what we need to do."

“I think regulatory sandboxes are very resource intensive [they can only] work with a small number of companies at a time, just because of the limits of the tool… they are only appealing to perhaps a certain type of cohort, maybe startup or smaller scale companies.”

Fragmentation and Limited Regulatory Awareness Undermine Progress: Lack of coordination between regulators, developers, funders, and the frontline healthcare providers (like the National Health Service in the UK) —combined with limited regulatory awareness among engineers—slows effective, safe AI deployment.

Section 3: Participatory Design & Oversight – Building Trust Through Inclusion

The roundtable’s final discussion focused on the role of participation in building legitimacy and trust in AI-enabled healthcare. Participants emphasised that current engagement practices are often tokenistic, narrow in representation, and structurally undervalued. To be meaningful, participation must be resourced, representative, and embedded across the full AI lifecycle from design and procurement to deployment and oversight.

Key Takeaways from This Discussion

- Participation must go beyond tokenism, involving patients, clinicians, and the public throughout the AI lifecycle.
- “Expert patients” and embedded clinicians do not reflect typical user perspectives; diverse and underrepresented voices must be included.
- Participation requires dedicated funding, skilled facilitation, and institutional support to be effective and sustainable.
- Post-deployment feedback and adaptive regulation are essential for learning, accountability, and public trust.

Main Discussion Points

Participation Must Be Continuous and Inclusive: Current engagement practices are often superficial, with input drawn from a narrow pool and limited to early or late phases. Effective participation must be cyclical, embedded across the AI lifecycle, and designed to reach underrepresented communities and typical frontline users.

“I think often with people and clinicians that are involved in the design of these technologies, they are not what I would consider the average clinician or the average patient, and they tend to be quite clued up. I do think it is possible to get a wider range of people. But it's often not the case that we do.”

Resourcing and Mandates are Essential: Participation is labour-intensive and often unfunded, placing unfair burdens on public bodies. Engagement must be incentivised or mandated in regulation and procurement, with shared funding responsibilities and investment in inclusive design skills.

"it's often seen as something that's nice to have. But we could actually provide the appropriate training materials to make sure that that is our routine… And then we are genuinely meeting the needs of the most vulnerable."

Post-Deployment Monitoring and Feedback Loops: AI regulation must extend beyond deployment to capture emerging risks, near misses, and real-world consequences. Scalable tools—including AI-supported feedback—can help, but must not replace human judgment or accountability.

"Thinking about some of the comments around impact assessment… I don't know how many teams or organisations then go back to update their Data Protection Impact Assessments, when actually, the majority of impacts are felt once the tool is released into the wild…"

Conclusion and Emerging Policy Recommendations

The roundtable surfaced a number of converging challenges and opportunities regarding the integration of AI in clinical decision-making. While participants represented a range of institutions and roles - from regulators and clinicians to researchers and policy professionals - common themes emerged across discussions. These themes point toward a set of policy directions that are both feasible and necessary to support the safe, effective, and trusted use of AI in healthcare.

Emerging Policy Recommendations

Mandate and Fund Meaningful Participation: Move beyond tokenistic engagement. Participation from clinicians, patients, carers, and underrepresented communities should be a regulated requirement, not an optional add-on, in the design, procurement, deployment, and evaluation of AI-enabled healthcare tools. Funding must reflect the time, training, and facilitation required to ensure participation is inclusive, representative, and sustained across the lifecycle.

Clarify Legal Responsibility for AI-Driven Decisions: Urgently establish nationally recognised guidance on liability in shared human-AI decision-making contexts. Fear of accountability gaps, particularly around clinician exposure in the event of AI-related errors, is stalling adoption. Regulators must define clear, consistent standards for how responsibility is shared across clinicians, developers, vendors, and NHS bodies.

Scale Regulatory Sandboxes and Broader Innovation Support: Expand regulatory sandboxes as part of a wider, joined-up innovation infrastructure. This includes shared virtual testing environments, coordinated procurement pathways, and consistent guidance to help navigate regulatory expectations. Given their resource intensity, sandboxes should be linked to broader, scalable support tools, and their insights must directly inform regulatory updates and standards.

Improve Workforce Readiness Through Systemic, Cross-Disciplinary Training: Embed training about AI, data governance, and accountability for all roles at different stages including clinical education, continuous professional development, and digital staff induction. Multidisciplinary roles should be introduced in deployment teams to bridge technical and clinical expertise. Engineers and developers must be equipped with context-specific knowledge of healthcare workflows, ethics, and patient safety imperatives.

Standardise Post-Deployment Monitoring and Human Oversight: Establish clear guidance and regulatory expectations for lifecycle oversight of AI systems, extending well beyond initial deployment. This should include scalable post-market surveillance, routine updating of impact and risk assessments (e.g. DPIAs), and transparency standards such as explainability of outputs. Oversight must be context-specific, risk adjusted, and proactive, supported by real-world feedback loops but grounded in human accountability.

2.3 Policy Roundtable - AI Governance and Trust in Climate Services

Why Climate Services need Responsible and Trustworthy AI Tools

The integration of AI into climate services promises major advances in forecasting, modelling, and climate risk planning. However, these benefits rely on climate scientists' willingness to adopt tools and institutional providers of climate services maintaining justified public trust. Despite the proliferation of ethical AI principles and regulatory discussions, operational AI governance frameworks remain underdeveloped. Regulation is a fragmented patchwork of national and regional strategies. For example, the UK's pro-innovation,[16] mission-driven approach coexists with the EU's risk-based framework,[17] creating potential barriers to international collaboration and data exchange that could enable AI use cases in climate services. With climate services playing a critical role in public risk communication, emergency response, and long-term resilience planning, these governance gaps are consequential.

A unique component of national climate service agencies is their dual responsibility to serve the public good while also providing specialised services to commercial sectors. Their external-facing services encompass a wide range of products, from public warnings for severe weather that help people "stay safe and thrive" to bespoke advisories for critical industries like aviation and energy.

As a "global centre of excellence in weather and climate science," the UK's Met Office is at the forefront of this technological revolution.[18] Its mandate is to deliver world-leading weather and climate prediction data and intelligence to government, businesses, and the public. The Met Office's strategic outcomes include maintaining the UK's position as a world-leading modeling center by remaining at the cutting edge of both physics- and ML-based modeling.[19]

This roundtable brought together different perspectives from the Met Office (see Appendix for an overview of participants) to discuss actionable steps that can help unlock responsible,

---

[16] AI Opportunities Action Plan, Department for Science, Innovation & Technology, accessed 13/05/2025
[17] Act, E. A. I. (2024). The EU Artificial Intelligence Act.
[18] https://www.metoffice.gov.uk/research/ai/met-office-and-ai.Accessed 17/07/2025.
[19] https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/foundation-science/data-science-framework-2022-2027.pdf. Accessed 17/07/2025.

safe deployment of these technologies at the organisational level and more broadly. It was convened in the wake of an announcement by the World Meteorological Organisation's (WMO) Executive Council of their Action Plan on Artificial Intelligence.[20] The Action Plan emphasised the importance of aligning AI applications in the climate domain with scientific and ethical standards, as well recognising the need to effectively balance between caution and innovation. These themes were central to the discussions between participants, which focussed on 2 broad topics:

- Transparency, accountability and public legitimacy
- Ethics, participatory design and oversight

The roundtable was delivered in partnership with the UK Climate Projections (UKCP) team from the Met Office, who are members of the HACID consortium. We used Miro (see Figure xx) and facilitated group discussions to gather individual and group-level opinions on a range of topics.

Section 1: Transparency, accountability and public legitimacy

A discussion of the key challenges or barriers to successfully embedding AI tools into the operational processes of climate services highlighted the following factors:

Verification and Validation: It is a challenge to verify and validate AI models because they lack the "granularity of understanding" that physics-based models have. The methods for testing traditional models do not necessarily apply to AI.

"As soon as you take more statistical approaches like machine learning or you take Deep learning techniques like large language models, you lose some of that granularity of understanding about what is driving the outputs of the model."

Uncertainty about Future Performance: Confidence in an AI model's performance in the present doesn't guarantee its effectiveness in the future. The "complex correlation structures" learned from historical data might not be applicable to a future climate with significantly changed factors.

"Our confidence in the present day isn't necessarily an indicator of what our confidence should be in the future. It might be for a physics-based model because we trust the physics under that range of conditions, but we don't necessarily trust relationships."

Public Perception and Trust: A recent survey by the Met Office revealed a wide diversity of views and a lack of public trust in AI forecasts. Only about 50% of people believe AI forecasts would be accurate, and confidence levels are even lower. This contrasts with traditional forecasts (e.g. Numerical Weather Prediction based models), where there is

[20] https://wmo.int/news/media-centre/wmo-faces-future-action-plan-artificial-intelligence. Accessed 17/07/2025.

strong agreement on their accuracy and high confidence in using them for important decisions.

“For machine learning, a much smaller… only about 50%, probably a bit less on average, believe those forecasts would be accurate. And there's a much wider spread across each of the different groups.”

The discussion then moved on to focus on how institutional structures might enable, rather than inhibit, the safe and consistent adoption of AI across different roles and contexts. The following factors were discussed:

Adaptable Quality Assurance (QA) Processes: A single, uniform QA process may not be suitable given the breadth of services offered by an organisation like the Met Office. The level of QA should be scaled to the maturity of the information and the type of decisions customers are making.

Integration of AI with Human Expertise: AI can be used to provide a "first guess" in a forecast, with human experts then intervening to add trust and confidence to the output. This approach leverages AI for efficiency while maintaining human oversight for critical services.

“We can add trust and confidence to the output of artificial intelligence models because we have got that expert intervention somewhere in the cycle.”

Defining the "Applicable Envelope": There is a need for new ways of testing AI models and specifying the envelope of conditions under which a model is suitable or applicable. This involves understanding the range of conditions where a machine learning model applies and acknowledging that its historical performance may not predict future performance.

“I think we also have to slightly break some of our preconceptions about how we do the testing we might require. We [probably] almost certainly do require different ways of testing and different ways of specifying the envelope under which a model is suitable or applicable under different conditions.”

The last part of the discussion in this first section focused on the role of institutions like the Met Office in demonstrating that AI-enabled climate services reflect values such as fairness, inclusivity, and consent. Participants raised the following points:

Responsibility for Usability and Trust: As a national provider, the Met Office's core responsibility is to ensure that any tool it releases is usable and that it provides the necessary support to build trust and enable people to use it. This responsibility extends to guiding stakeholders on how to appropriately and effectively use other available tools, not just their own.

"The responsibility is on making things usable and as a National Met provider, just that feels absolutely core to what we do - that we need to take responsibility for any tool we put out."

Addressing Divergent Concerns: Internal survey results from research by the Met Office,[21] show different groups have different concerns and require different information to understand and use AI-enabled services effectively. Addressing these diverse needs is a significant but important task, especially for critical services like weather warnings.

"There will be different evidence required by different groups, [..] different information that will be required by different groups to help them understand what we're doing and how they can use the information effectively."

Leadership in a Novel Field: The Met Office has a role to show leadership in the space of AI assurance and advice. Since climate service standards are still a novel concept, AI can serve as a "useful platform" or "vehicle" to initiate discussions and build more robust QA processes and trustworthy systems.

"The Met Office needs to show some leadership here. Amongst all its stakeholding communities and government and international net service communities and saying, yeah, we are not on top of it right now."

Section 2: Ethics, participatory design and oversight

In the second section of the discussion, participants were asked what roles should different actors - including government, experts, affected communities, civil society, and the wider public - play in the design, oversight, and evaluation of AI in climate services? How does this vary between different climate services contexts?

Government's Reliance on Met Office: The UK government relies on the Met Office to provide assurance for weather and climate information regardless of the technique used. However, the Met Office must also adhere to international standards, such as those from the World Meteorological Organization (WMO).

"The UK government relies on the Met Office to provide assurance of weather and climate information irrespective of what technique has been used to generate that."

Varying Accountability: Public accountability varies depending on the service's audience. For services to a narrow group like defense stakeholders, there may be little public knowledge, but the decisions are critical, requiring a high level of verification and QA through a specific chain of command.

"The Met Office is a government organisation and therefore publicly accountable. But I think for many of the things we're doing here, they are sort of projects or services focused on a relatively narrow group of people or organisations."

Cultural Differences in Trust: Global surveys show significant cultural differences in public trust in AI. The UK has low trust, while countries like Nigeria and India have high trust. This

[21] One of the roundtable participants had just completed a survey study measuring attitudes to AI and shared some of the emerging results in the course of conversations. (Publication of these results is in preparation.)

suggests that the types of participation and assurance processes needed may differ across countries.

"When you look at global surveys of trust in AI... The UK is probably in the lower third in terms of kind of trust in AI. Finland I think has the lowest trust and countries like Nigeria I think it's like 90%. Plus, people's trust in AI in India, similarly, is kind of quite high."[22]

In terms of practical tools that can help balance the tensions between accessibility, innovation and safety, participants suggested the following:

Argument-Based Assurance: participants mentioned testing the "Turing Framework trustworthy and ethical assurance". This framework requires developers to write down claims about a service (e.g., that it is fair or explainable) and provide evidence to back up those claims. This process allows for building QA procedures and agreeing on claims and evidence with users.

Learning from Other Sectors: The Met Office should look "outside of our field and learn from other fields" such as the medical or financial sectors that have been grappling with similar challenges. This includes learning from both good and bad practices.

Iterative Process: People also discussed that the QA processes for AI will "evolve" and they may eventually develop standard questions or claims that need to be tested each time.

When asked about what forms of participation, consultation, or shared governance are needed to ensure that AI tools used in climate services are publicly accountable, participants suggested the following:

Whole-Process Stakeholder Involvement: Stakeholders should be involved throughout the entire process of developing AI and machine learning-based services, not just at the end. This includes working with a representative stakeholder group from the start to identify the problem and design a tool that is "fit for purpose" and accessible to the user community.

"involve the stakeholders and [you know, with] different sections of society, not just in the testing of the final product but in the design of any AI or machine learning based services."[23]

Beyond External Trust Building: Building trust is not just an external challenge with the public; it is also an internal one. It is important to bring the scientific community along with new AI-enabled processes, in addition to the user and customer communities.

"I think it's easy to assume that trust is just external… and I don't think it is. I think we've got to bring a scientific community with us as well as a user and customer community."

---

[22] Participant was referring to results from Edelman's Trust Barometer, 2025 (https://www.edelman.com/sites/g/files/aatuss191/files/2025-02/2025%20Edelman%20Trust%20Barometer_Insights%20Technology%20Sector_FINAL.pdf. Accessed 29 August 2025)
[23] Quote edited for clarity with use of square brackets.

Emerging Recommendations

Based on the roundtable discussions, we recommend the following actions to enable responsible and effective use of AI in climate services:

For National Meteorological Organisations (like the UK Met Office)

- **Lead the way in robust AI assurance:** National meteorological organisations should take a leadership role in advising on the use of AI in climate services for its stakeholders, including government and international communities. Instead of simply adopting AI, it should use its unique position as a government agency to show how to build trust and ensure safety, particularly for public-facing services and other public-interest applications such as defence.
- **Adapt quality assurance processes:** Existing quality assurance (QA) processes for climate services should evolve to specifically address the unique challenges of AI and machine learning. This includes developing new methods for testing and validating models, particularly for how they might perform in future climates, and creating standard claims that must be verified for each project. The level of QA should be flexible and scaled to the maturity of the information and the risks associated with its use.
- **Invest in human oversight:** Any use of AI by climate service providers should adopt a "human-in-the-loop" approach for high-risk applications, such as severe weather warnings. AI can be used to provide a "first guess," which is then checked and verified by expert meteorologists, adding a layer of trust and accountability.
- **Foster internal and external trust:** Climate service providers should not only focus on building trust with external users but also with its own scientific community. This can be achieved by transparently documenting the techniques used and engaging scientists in the process of building a body of knowledge around AI.

For National Level Regulators

- **Establish a risk-based framework:** Working closely with public and commercial climate service providers, regulators in the UK and other countries should create a clear hierarchy of risk for AI applications in climate services, defining what constitutes high, medium, and low-risk uses. This framework should specify the necessary assurance mechanisms for each category.
- **Create clear disclaimers:** Regulators should encourage the use of "validity narratives" for AI-enabled climate services. These disclaimers would clearly state where a model’s accountability extends and where it stops, providing users with a clear understanding of its limitations and appropriate use.
- **Mandate argument-based assurance:** Regulators could mandate a governance framework like the Turing Framework or recognised responsible AI standards like ISO 42001, which requires organisations to make explicit claims about an AI service's fairness, inclusivity, and explainability and provide evidence to back those claims up. This provides a transparent and verifiable record of the development process.

For International Actors (like the WMO)

- **Develop New International Standards:** The WMO should develop a new set of international standards and metrics specifically for the evaluation of AI models in weather and climate applications. These new metrics are necessary because traditional verification and validation methods for physics-based models do not fully apply to AI.
- **Coordinate global standards:** The WMO should act as a central hub for coordinating a global consensus on AI governance in the sector. This is crucial as public trust in AI varies significantly across different countries and cultures, affecting the demand for assurance.
- **Learn from other sectors:** International organisations should look beyond the climate sector to learn from other fields that are grappling with similar AI governance challenges, such as healthcare or finance. This can provide valuable insights into both good and bad practices.

For Research Funders

- **Prioritise foundational research into the limits of AI:** Funders should invest in foundational research to better define and document the "envelope" of conditions under which machine learning models for climate services are applicable.
- **Support participatory design:** Funders should prioritise projects that involve stakeholders and communities throughout the entire development lifecycle of AI tools, from initial design to delivery and use. This ensures that the tools are fit for purpose and are representative of the communities they are meant to serve.
- **Support cross-institutional collaboration:** Research funders can support projects that bridge the gap between different parts of organisations, different sectors and different geographies to address the multidisciplinary nature of AI governance and current gaps. This can help bring together diverse expertise from climate scientists, data scientists, and stakeholders on the ground, including communities affected by climate risks.

3. Alternative Use Cases: Extending the HACID Approach to other domains

HACID defines a methodology for decision support that can be generalised to different application domains beyond health diagnostics and climate services, where expert knowledge is vast and difficult to integrate, requiring at the same time the contribution from human experts and the data processing abilities of AI systems. We used a structured design process to identify and prioritise relevant domains. The shortlist included:

1. Disaster response and risk reduction,
2. Water regulation and management,
3. Asylum determination management,
4. Energy systems.

These domains were selected due to the inherent complexity and uncertainty of decision-making involved, the dynamic nature of expertise and data on these topics and their perceived value to policy makers as a proxy for potential future impact.

For two of the shortlisted domains, Energy Systems and Water regulation and management, we were able to identify partner organisations who could provide relevant domain expertise to help us a) scope out and identify a use case; b) develop an early-stage blueprint for a HACID-inspired tool, and c) provide feedback to refine and assess the desirability/viability of the blueprint. In the next sections, we describe the design process and resulting insights, and present the associated design artefacts. We finish with concluding reflections about the value of exploiting a HACID approach in other domains.

3.1 Alternative HACID Use Case 1: Energy Sector

Design partner(s)

The exploration of this domain was completed in collaboration with students from the Masters in European Design (MEDes) at The Glasgow School of Art, who led on research and design activities, with oversight from Nesta's Centre for Collective Intelligence. Additional domain expertise was provided by Nesta's Sustainable Future team whose mission is to reduce carbon emissions by 30% by 2030.

Why this domain is suitable for a HACID decision support system

Navigating the future of energy systems requires making complex decisions with immense uncertainty. A decision-support tool that combines AI and crowdsourced human expertise can be highly relevant for this task. Generative AI and ML can analyze vast amounts of traditional data (e.g., weather patterns, grid performance) alongside non-traditional sources like qualitative perceptions of local residents and real-time behavioral data. This approach aligns with calls from organizations like the International Energy Agency (IEA) for the integration of social and behavioral insights into energy system modeling to enhance accuracy and public acceptance (IEA, 2023). Furthermore, this combination addresses the challenges of fast-paced technological change and emerging information in the energy sector, a point underscored by numerous studies on energy transition (e.g., Sovacool et al., 2021). By incorporating crowdsourced human expertise, a HACID-approach could continuously integrate the latest knowledge and refine its models, making it an agile system that adapts to changing conditions and helps optimize the location and operation of energy sources within a community, city, or region. There have also been calls for "human-in-the-loop" models to effectively manage the uncertainty inherent in energy transitions and to help facilitate uptake of sustainable behaviours.

Three additional factors shaped our exploration of how a HACID approach could be used to support decisions about energy:

- **Democratisation and Community-led Energy Systems**
  There is a growing number of examples of decentralised, cooperative energy models where local communities participate in decision-making and resource management.
- **Integration of Ecological Knowledge and Regenerative Systems**
  A sustainable energy transition requires incorporating bioregional knowledge and

regenerative energy sources, ensuring long-term ecological sustainability while strengthening local resilience.

Government-led initiatives, in collaboration with energy cooperatives, will be key to Scotland's ability to navigate this energy transition, reinforcing the shift towards a more sustainable, community-driven energy landscape. This context provided the opportunity to extend the HACID approach beyond the support of decision-making by professionals.

Overview of Design Process

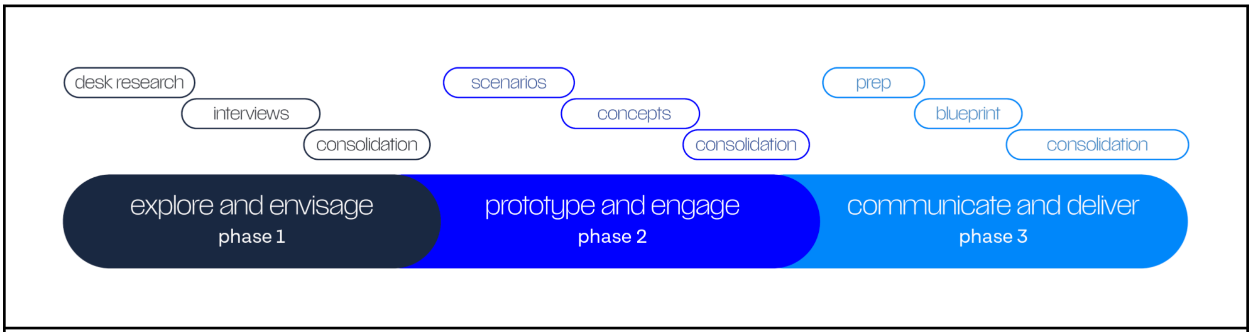


Figure 3: An overview of the research and design process.

To develop and design this blueprint and future system, we followed a three-phase design approach. Each phase corresponded to a particular set of research and design methods to develop a deeper understanding of the context. This evidence base was used to create the proposed energy blueprint and six prototypes for concepts or touchpoints that support the implementation of the blueprint.

Phase 1: Explore and Envisage

Core activities: Desk research using STEEPLE cards, Interviews with domain experts, Casual Layered Analysis workshop.

STEEPLE cards offer a **structured approach to domain research** (see Figure XX). They help with exploring the context from different perspectives using the framework of categorising insights into Social, Technological, Economic, Ethical, Educational, Political, Legal, Infrastructural and Ecological themes. This approach was used to gain a deeper understanding of the opportunities and challenges for current and future Energy Systems. The emerging evidence base around energy systems technology, policy priorities and public opinion were also tagged according to their relevance at the global, national or local scale. All insights and evidence gaps were tested for accuracy and relevance through **interviews with domain experts.**

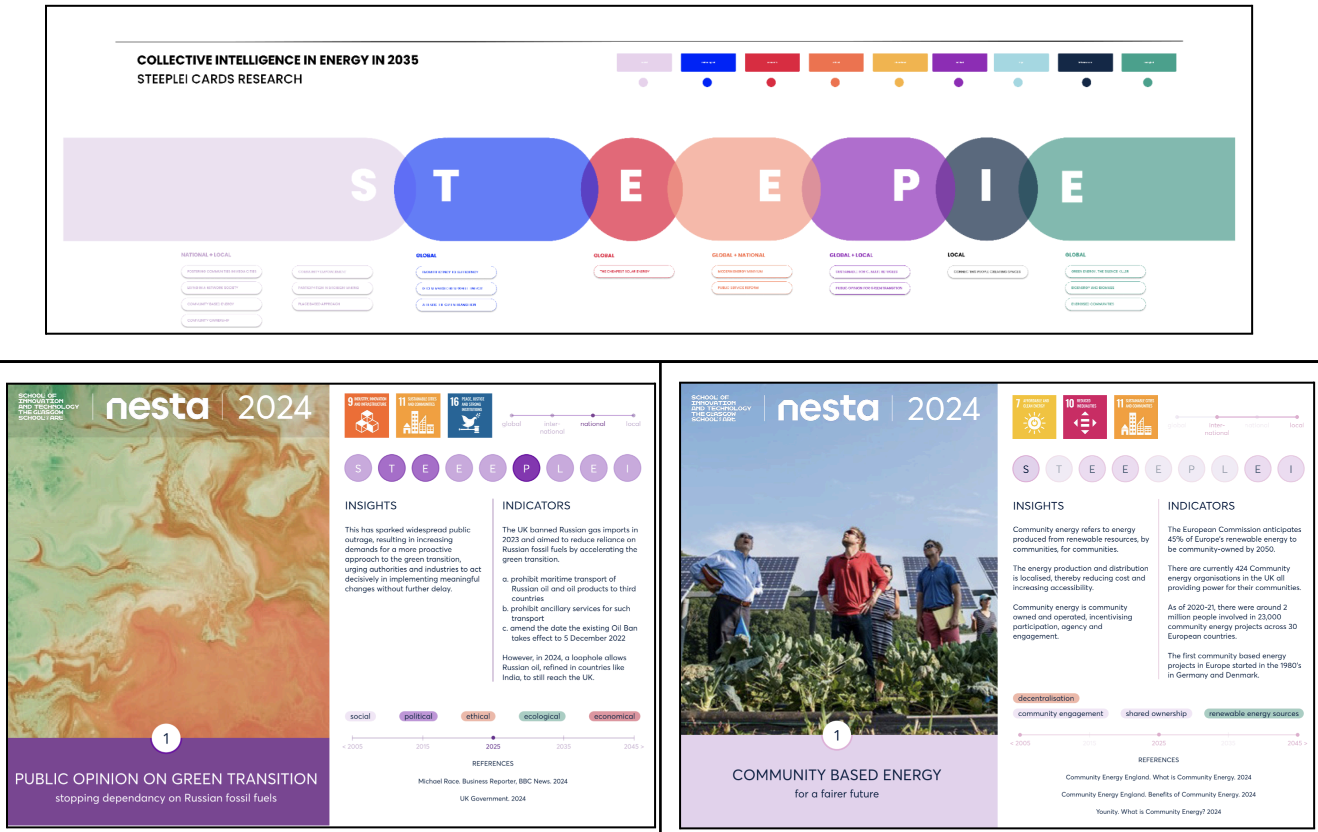


Figure 4: The top figure shows an overview of all emerging insights categorised according to the STEEPLE framework. The lower figure shows two example STEEPLE cards with insights related to Public Opinion on Green Transitions and Community Based Energy. High resolution images are provided as a separate attachment.

Key insights from desk research and interviews

The key insights from desk research and interviews were grouped into four key design principles, collaboratively identified by the students and Nesta. These key themes guided subsequent design activities.

1. **Equitable Access:** the democratisation of resource management via community-led decision-making ensures universal energy access supported by shared root systems.

"The energy future of the world is not only dependent on one technology. The advantage of solar energy is that it enables more active participation from people by bringing power into their own hands. The decentralisation and democratisation of energy systems could help nudge the conversation closer to the Paris Agreement targets."[24]

24 https://journal.urbantranscripts.org/article/decentralising-and-democratising-energy-guest-editor-maria-tzika-talks-to-neel-tamhane-solar-strategy-lead-from-space10/

2. **Ecological Knowledge:** prioritising situated knowledge and community reciprocity to work towards a system led by 100% regenerative sources.

3. **Energy Literacy:** empowering individuals through energy knowledge requires effectively communicating how energy is produced and used.

4. **Community and individual value:** developing new models for personal responsibility and community participation via shared energy initiatives.

The Glasgow City Council voted to declare a climate emergency in May 2019, stating that: "Local communities should be able to invest in low carbon solutions through mechanisms such as cooperative business models. There is a good opportunity to develop partnerships which will enable residents to support renewable energy in the city and to get local benefit from these resources." [25]

Causal layered analysis (CLA) is a futures methodology that helps to articulate new scenarios or technologies for the future by examining underlying assumptions around a topic that might influence the current status quo or facilitate the uptake of new ideas. It supports the analysis of a particular domain through 4 layers: the common understanding of an issue (surface layer, e.g., newspaper headlines), the causes creating and sustaining the situation (systems layers), the different perspectives shaping it (actors layer) and the underlying stories that might feed it (myths/metaphors layer). Using a CLA approach, we developed a vision of the future energy system in Scotland for the year 2035 (Figures 4 and 5). Rather than exploring speculative possibilities, we concentrated on identifying the most probable outcomes based on the research insights.

25 https://carboncopy.eco/initiatives/glasgow-community-energy?gad_source=1&gclid=Cj0KCQjw3vO3BhCqARIsAEWblcAbQnsTZkw4m_x2qH_tskqaE5F_dEbHfiveYxTz42iv_L9bP93YoEsaAlOcEALw_wcB

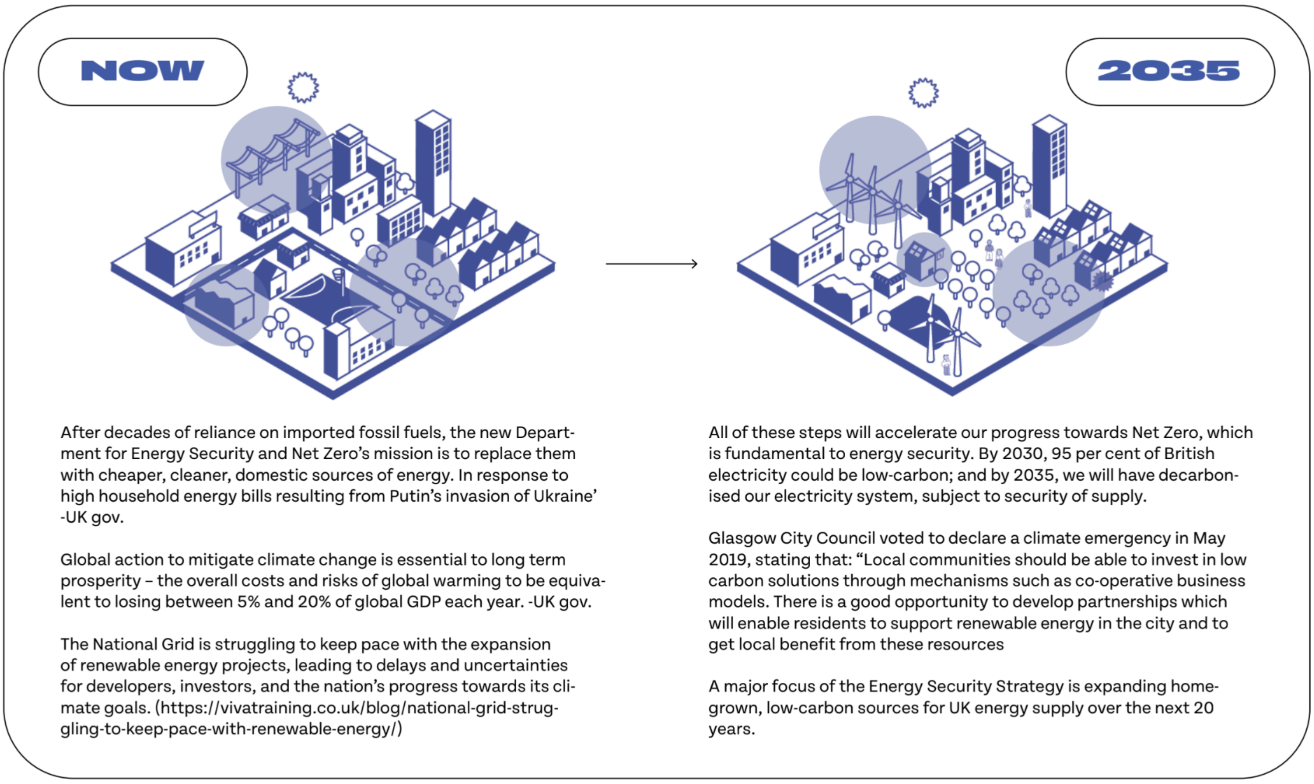


Figure 5: Overview of the current (Now) and future World Vision (2035) of the energy system in Scotland, developed through the causal layered analysis approach. High resolution images are provided as a separate attachment.

Vision for 2035: Building the Future Today

Dennistoun, a neighbourhood located in the east end of Glasgow, was chosen as a case study location and context to anchor the project. Dennistoun is a well-known area with a rich legacy of collective action and community-driven initiatives. One prominent example is the Alexandra Park Community Centre, a local engagement and collaboration hub. The Dennistoun cooperative serves as a model of success due to robust community engagement and support. This highlights the feasibility of Dennistoun as a setting for renewable energy initiatives led by local communities.

Vision statement: *"By 2035, local energy cooperatives will play a critical role in meeting the diverse energy needs of different regions, further cementing the transition to a sustainable, community-driven energy future."*

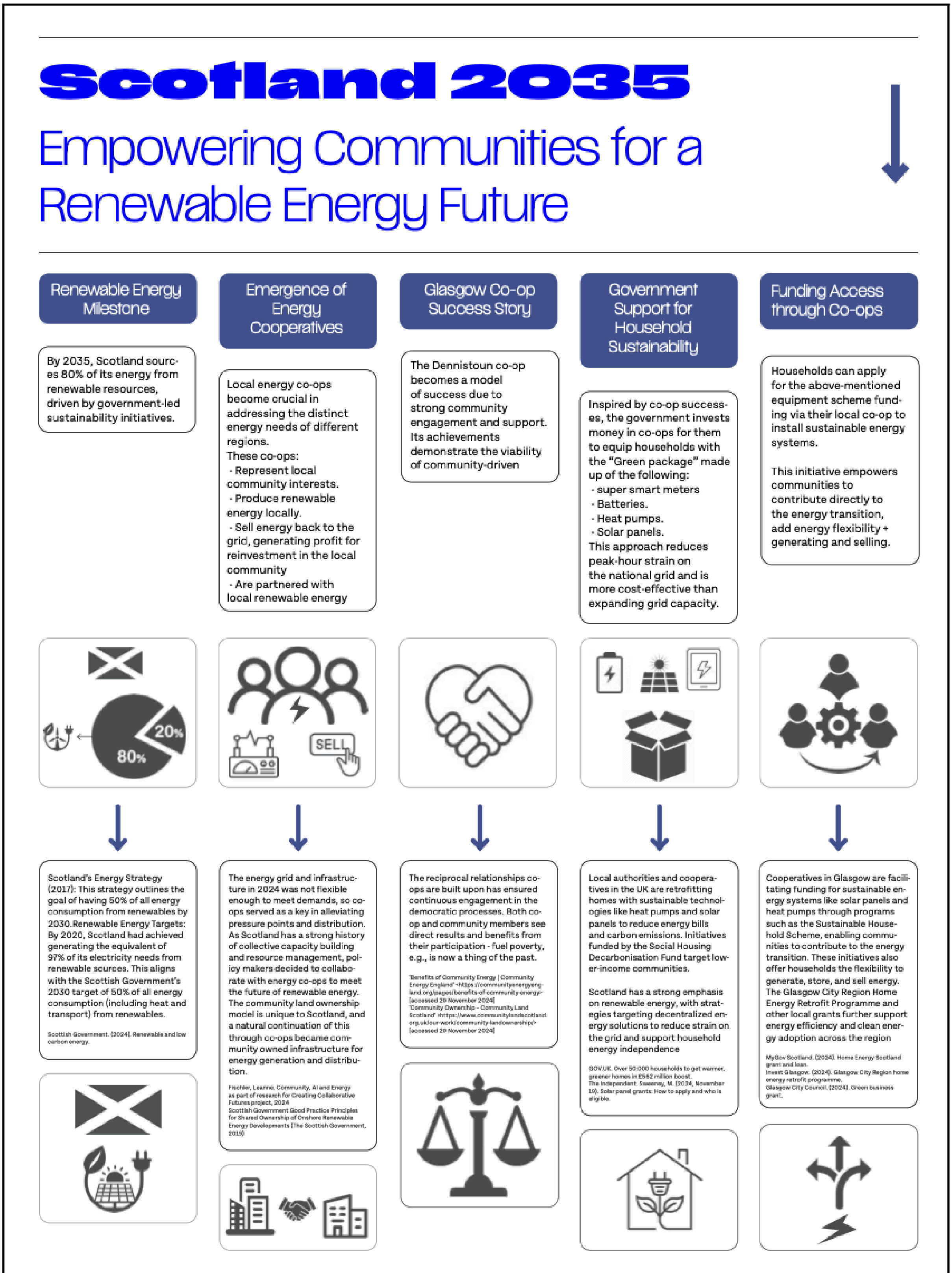
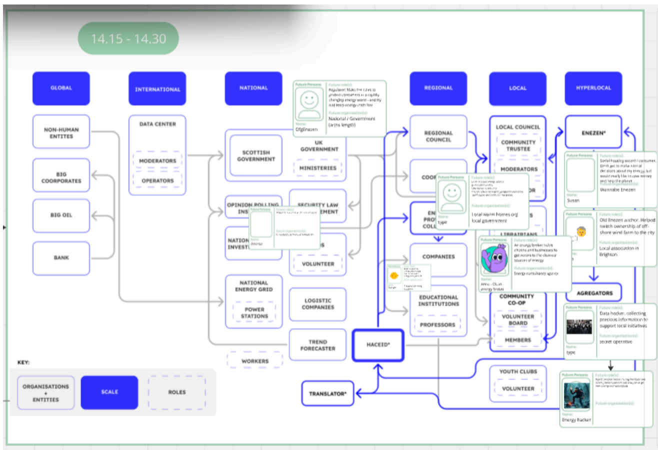


Figure 6: Further detail of the 2035 World Vision. High resolution images are provided as a separate attachment.

Phase 2: Prototype and Engage

Core activities: Concept co-creation & System mapping workshops, Prototype development.



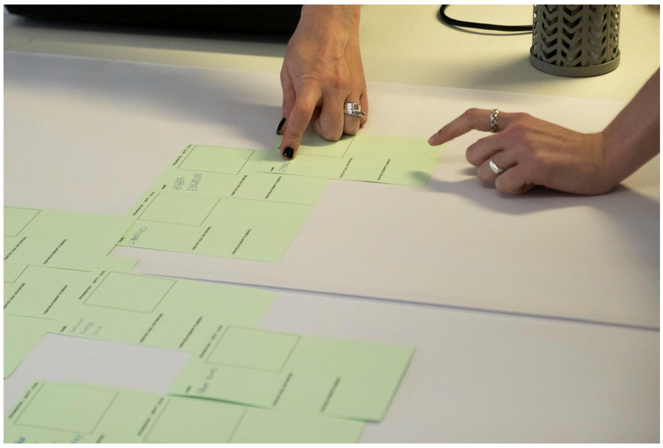


Figure 7: Left: Concept co-creation workshop (screenshot) from collaborative feedback session from Sustainable Future team and HACID consortium members. Right: Systems mapping workshop.

During this second phase, we designed individual responses to the opportunity defined within the Future World Vision. We tested the viability and feasibility of the emerging design concepts in a **co-creation workshop** with experts from the Sustainable Future team and technical experts from the HACID consortium. We followed this with a **systems mapping workshop** to gain a clearer understanding of the main stakeholders, technical components and other entities to consider in the local (Scottish) energy ecosystem, and how these elements interacted with one another. These workshops were essential in **testing and refining the 5 different touchpoints** (Figure 8, further detail on the touchpoints is provided in the Appendix) for individuals to interact with the emerging HACID system and to gather the foundational building blocks for the blueprint design in the next phase.

Key insights from the Develop phase

- **Community-driven decision-making requires both technology and human facilitation.** AI can support collective intelligence, but trust, accessibility, and local lived experiences must be central to ensure meaningful participation.
- **Energy literacy is a prerequisite for engagement.** Without clear, interactive, and contextualized information, individuals cannot make informed decisions or contribute effectively to sustainable energy initiatives.
- Smart technology alone won't drive change—**behavioural and social incentives are essential.** Data transparency, community ownership, and ethical considerations must be prioritized to ensure adoption and long-term impact.

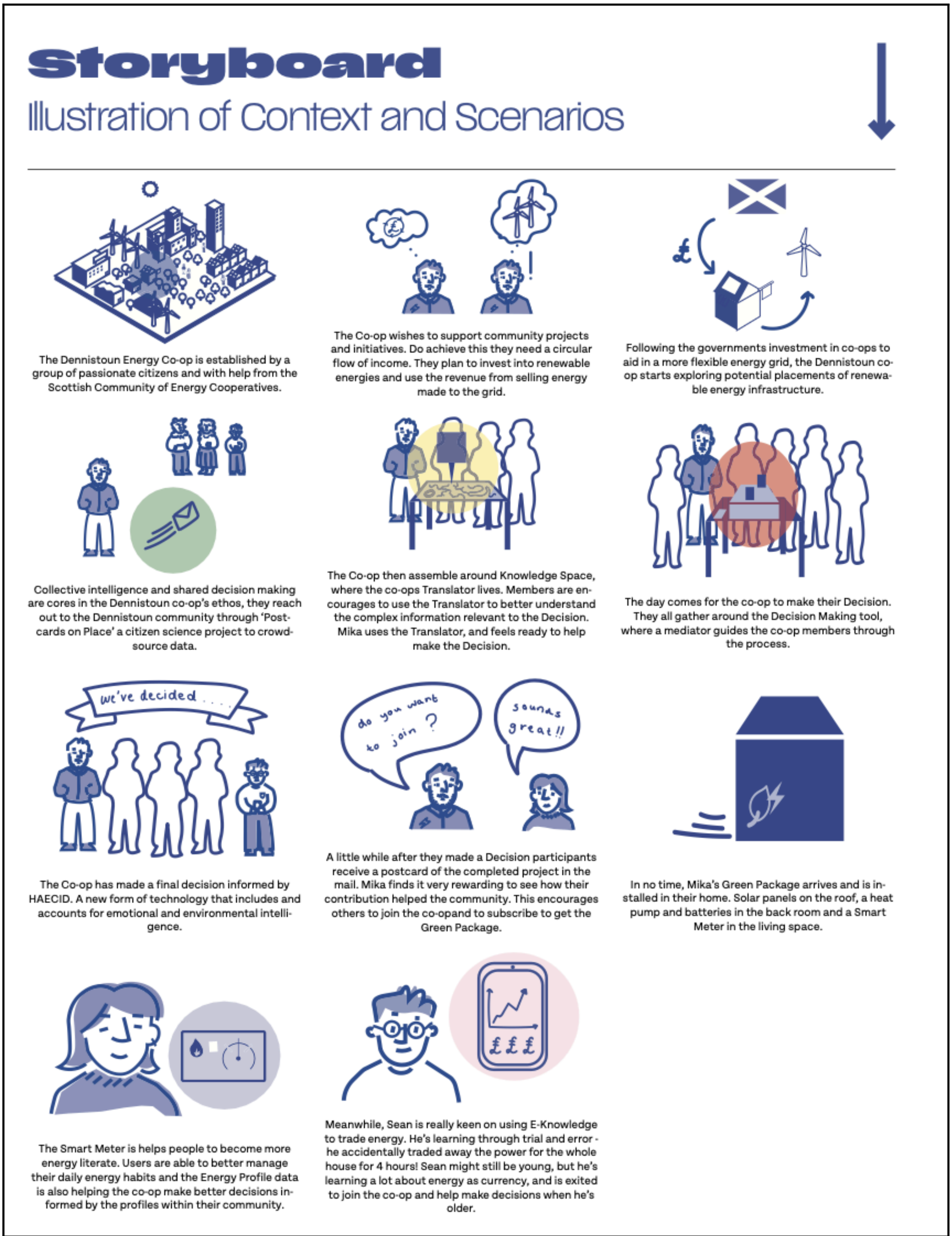


Figure 8: Scenarios illustrating the 5 concepts/touchpoint prototypes (coloured circles). High resolution images are provided as a separate attachment.

Phase 3: Communicate and Deliver

Core activities: Prototype showcase, Developing the blueprint design, Delivery of scenario and final blueprint.

We organised a **prototype showcase**, with physical artefacts representing the 5 different touchpoints developed in the previous Phase. During this showcase, HACID consortium members and experts in energy systems (Sustainable Futures team) interacted with the prototypes and provided feedback on relevant datasets, possible integration algorithms and incentives for people to contribute their data through the various interfaces.

Image 2 : Showcase of the 5 prototype touchpoints and the v1 HACID blueprint.

In the final phase of the project, we identified a more specific decision-making scenario to map out the HACID blueprint:

'The Dennistoun community energy co-op has a community fund looking to invest in wind turbines but needs support on where to place them (within Dennistoun, Glasgow)'.

The HACID system will aid in deciding the placement of energy infrastructure by aggregating and presenting relevant data and ranked solutions at different stages of the process. The supporting systems of HACID facilitate this process, which includes generating case and domain knowledge graphs. The 5 prototype touchpoints developed in the previous phase work as facilitators, passive and active data gatherers, educators, and translators.

Blueprint overview and scenario, with prototype touchpoints

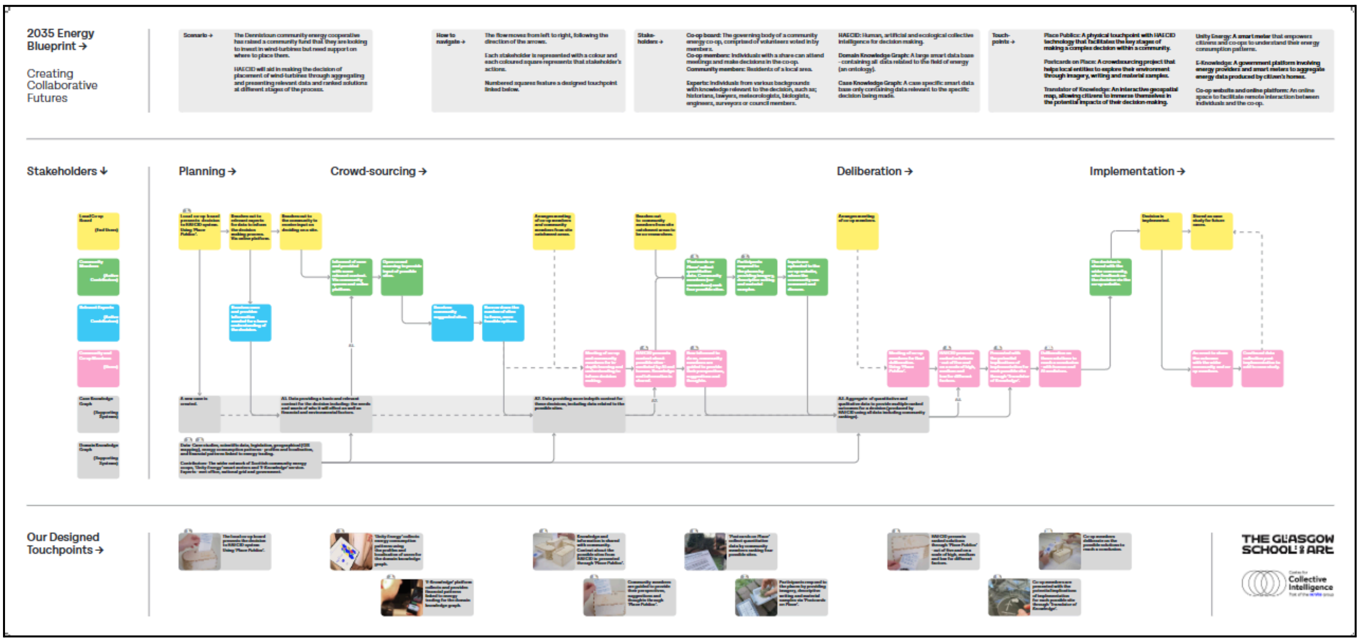


Figure 9: HACID Blueprint for Scottish Community Energy Cooperatives. High-resolution image is provided as a separate attachment.

The blueprint illustrates a future energy system that exists in 2035. It aims to implement a HACID-based approach to support decision-making for local Scottish community energy cooperatives.

Multiple stakeholders engage with the system throughout the process of reaching an informed and equitable decision via the designed touchpoints. These stakeholders include:

- **End users:** the cooperative board that initiates and implements the decision.
- **Contributors:** relevant experts and community members who actively provide both quantitative and qualitative data,
- **Users:** co-op members who make the decisions themselves.

Step by step scenario for the HACID energy systems blueprint
The blueprint illustrates how a future HACID decision support system would function. It maps the stakeholders involved, the datasets that are contributed by different actors and the technical infrastructure that is required (including knowledge graphs and algorithms). Below, we describe a step-by-step scenario for how the tool might be used from the perspective of someone living in a community with an existing energy co-op.

Context	You live within a small neighbourhood and discover that your community is planning on investing money into new renewable energy infrastructure for the residents. This project is managed by a local community energy co-op.
Planning Phase	1. The local co-op board gathers and raises the decision, through a deliberation tool of, ‘What is the best placement of new renewable energy infrastructure by and for the local community energy co-operative?’ - and so a case is created in HACID 2. Relevant experts contribute by narrowing down to fewer, more feasible sites through an online platform. 3. The co-op reaches out to the local community via open crowdsourcing on an online platform through which you can contribute possible sites.
Crowdsourcing	4. The co-op and broader community members gather in a shared space to learn and contribute via physical cards. 5. The catchment area is formed around a final selection of locations. Local residents like you are invited to be a co-researcher in the project: providing quantitative and qualitative data by ranking location sites based on preferences and material samples. 6. You meet other contributors to share and discuss thoughts with others within the catchment area.
Deliberation	7. A members-only co-op meeting is organised where they deliberate on ranked solutions presented via physical cards by HACID. This is facilitated by a translator that visualises the impacts of each solution.
Implementation phase	8. A decision is made and fed back to individuals and the broader community for feedback before implementation. 9. An event is organised to gather the wider community where everyone celebrates and marks the occasion. 10. The project is saved as a case study and shared with the wider co-op network for future reference.

Conclusion on HACID for Energy Systems

HACID provided a unique starting point to explore energy transitions and decisions being made about energy systems at different levels, from the individual and community-level, through to regional and global trends. The resulting blueprint challenges existing power dynamics of the traditional top-down flow of knowledge and decision-making. The tool acts as an aggregator of non-traditional data and local knowledge, such as experiential qualitative citizen reports, with quantitative data, such as individual-level energy use. Each component of the system was designed to enhance community-level understanding of energy and empower bottom-up decisions.

3.2 Alternative HACID Use Case 2: Water Regulation

Design partner(s)

The exploration of this domain was completed in collaboration with Ofwat, the Water Services Regulation Authority responsible for economic regulation of the privatised water and sewerage industry in England. The aim of the collaboration with Ofwat was to inspire policy audiences to explore the future of water management and regulation and how hybrid systems might support regulators to make better decisions around water management.

Why this domain is suitable for a HACID decision support system

Hybrid systems aim to leverage the best of human and artificial intelligence.They combine the power of technology to ingest and analyse large amounts of data with human expertise such as contextual and tacit knowledge. The UK — along with the global water management industry — is under mounting pressure, facing increasing challenges and uncertainty.[26] In the UK, water regulators are calling for greater investment and improved customer service, in response to ongoing infrastructure issues such as leakage and sewage pollution.[27] At the same time, evolving technologies and interconnected systems are driving rising public expectations around how services should be delivered.[28]

This collaboration was part of a broader effort to find out how HACID could help water regulators make more effective, transparent, and inclusive decisions, particularly where data is evolving rapidly and expertise is distributed. The HACID design workshop aimed to explore how AI and human expertise can be combined to enhance water regulation by mapping diverse data inputs, expertise sources, and open-ended regulatory questions. This work also aligned with the UK Government's pro-innovation AI Action Plan that encouraged regulators and other public agencies to identify new ways for AI to improve delivery of public services. The workshop was an opportunity for Ofwat to explore early strategic priorities for AI implementation and expand their thinking on how to best combine existing resources for complex decisions.

[26] https://www.nesta.org.uk/report/citizen-powered-water-management/
[27] https://www.gov.uk/government/publications/independent-commission-on-the-water-sector-regulatory-system-terms-of-reference/independent-commission-on-the-water-sector-regulatory-system-terms-of-reference
[28] https://www.ofwat.gov.uk/publication/trust-and-perceptions-peoples-views-on-the-water-sector-full-report/

Overview of the Design Process

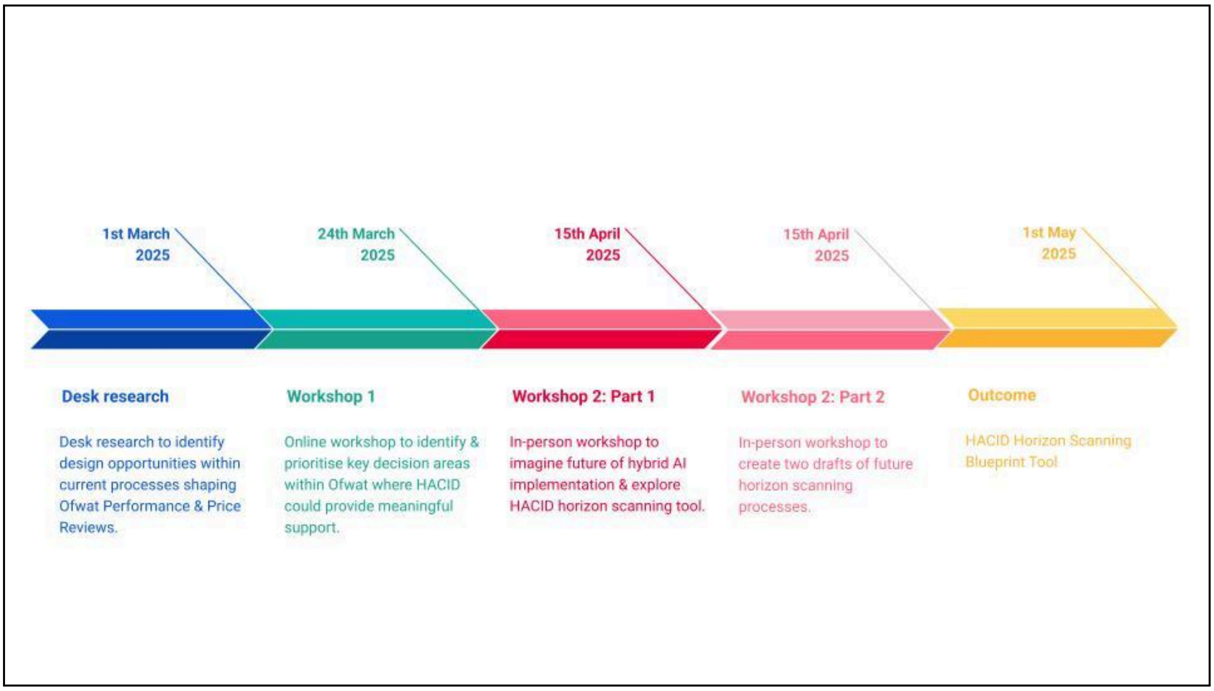


Figure 10: Overview of the design process for the alternative HACID use case 2: Water Regulation

We drew on elements of speculative design, a particularly useful approach in complex and evolving domains like water regulation. Speculative design allows stakeholders to explore possible, plausible, and even provocative futures in a structured way. By stepping outside the constraints of current systems and assumptions, it opens space for imagining how decision-making processes could adapt—or be reimagined—in response to emerging technologies, shifting environmental pressures, or evolving public expectations.

The outcome is a co-designed blueprint for a horizon scanning process that captures high-level, speculative propositions on how decision-making could evolve. The blueprint is similar to a map combined with an actionable plan that addresses a specific decision-making process to enhance service quality and customer satisfaction. It can vary in terms of the level of detail and the nature of design being more realistic or speculative. The blueprint sets out how multiple, including non-traditional, data sources and human expertise can be used in the water regulation decision making process.

What we did and key insights

Desk research

In collaboration with Ofwat, we gained valuable insights into the processes shaping the Performance and Price Reviews - two core internal Ofwat functions - and identified several opportunities for hybrid intelligence to support decision making both within Ofwat and across the water management sector more broadly. Key observations gathered from the desk research included:

- the need for more comprehensive data collection and water quality monitoring, supported by flexible, agile, and collaborative data practices;

- the need for better coordination in reporting and data sharing across regulatory bodies;
- the need for greater visibility on the local impacts of water companies, leading to stronger accountability and more effective local stewardship;
- the need for broader co-creation of water management strategies, involving a wide range of stakeholders; and
- the need to create new ways for customers to protect their interests.

These observations were gathered during discovery research prior to this work, which fed into the exploration of the use of HACID in alternative domains.

Workshop 1 - gathering deeper insights and identifying potential drivers of change

In the first online 2hr long collaborative workshop with Ofwat experts, we identified and prioritised potential drivers of change - processes within the water management domain that HACID could support - building on previously identified opportunities. We identified the following processes:

1. General horizon scanning and risk identification
2. Protecting consumer interests, particularly during performance assessments and price reviews
3. Monitoring company performance and service quality
4. Evaluating the environmental performance of water companies

We also explored several critical questions:

- Why is this tool needed? What are the main pain points?
- What data sources are currently available?
- Where are the data gaps?
- Who are the main stakeholders involved, and who else holds relevant expertise?

Through this process, we learned that the most impactful decision-making areas, both for Ofwat and the wider sector, are the **Price Review process** and **general horizon scanning and risk identification**. The latter, in particular, shows strong potential to enable **proactive** rather than reactive responses to issues like environmental harm, supports the adaptation of **emerging and evolving regulatory responsibilities, and** addresses **dependencies on external bodies** beyond Ofwat's direct oversight (such as the Environment Agency). Based on these insights, we decided to **develop a blueprint for a general horizon scanning and risk identification process**.

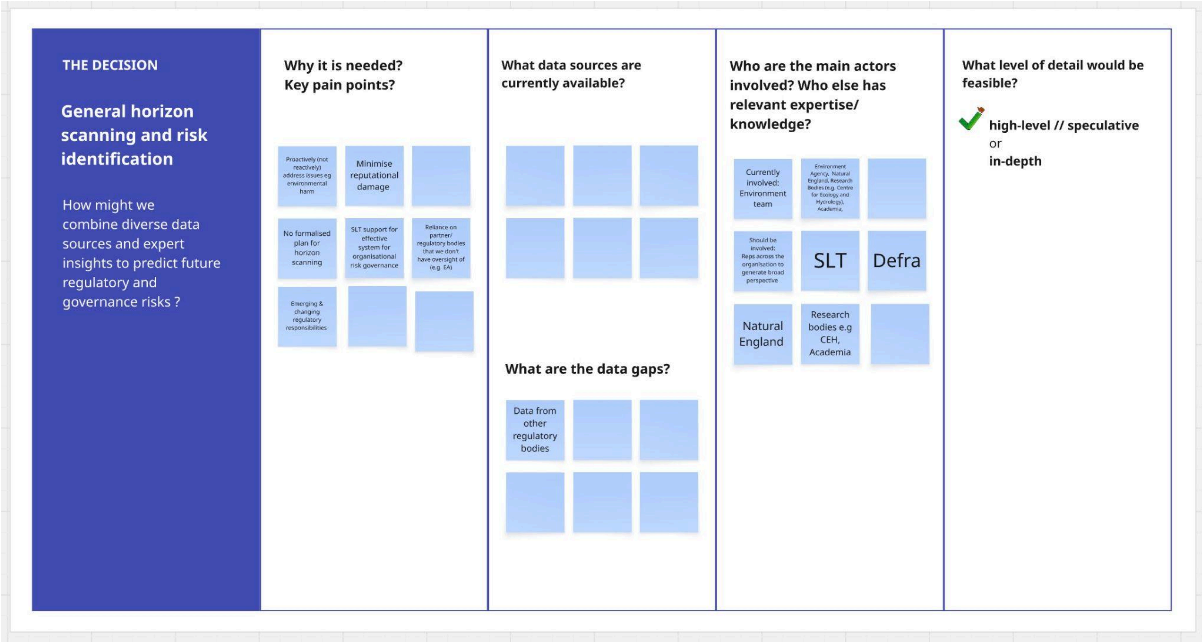


Figure 11: An example of one of the five design canvases, showcasing the questions presented to Ofwat participants during the first workshop.

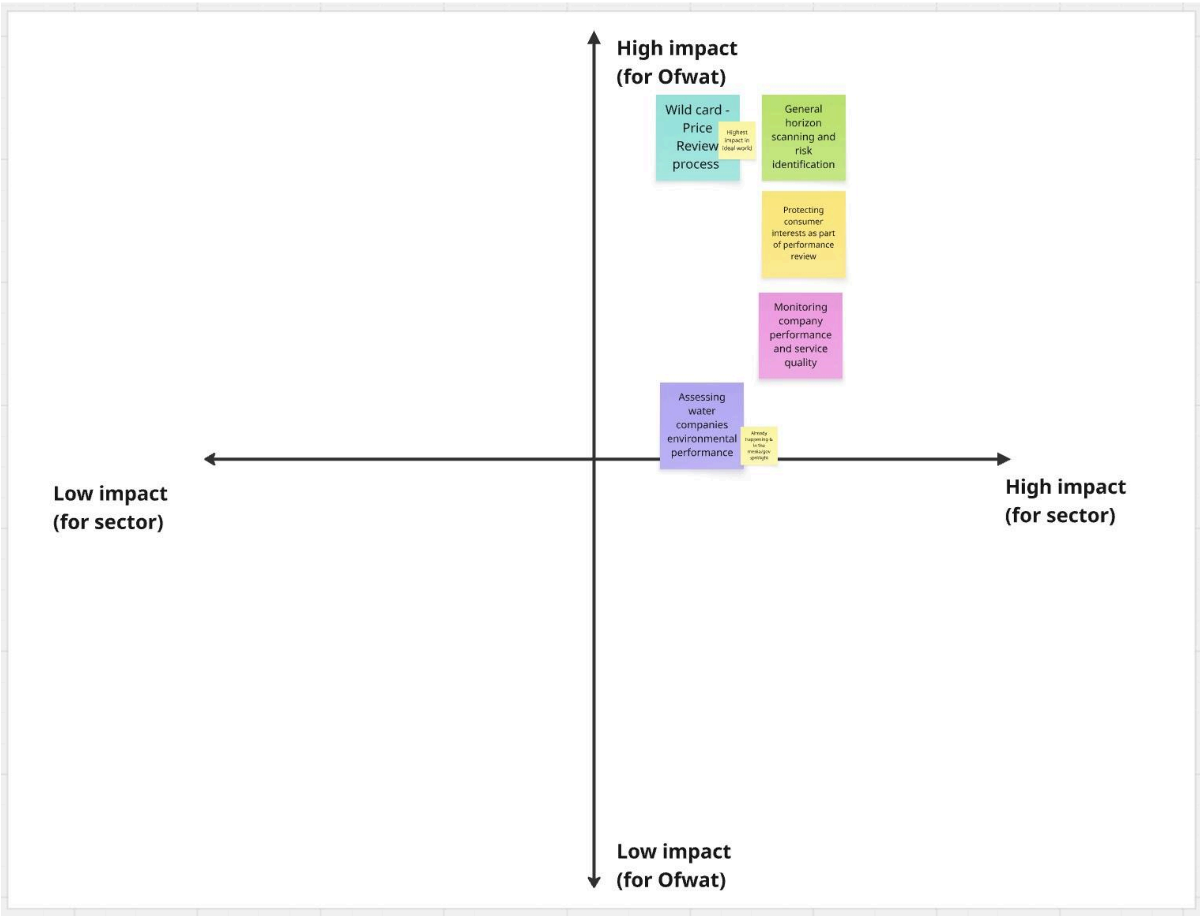


Figure 12: An impact prioritisation activity led to the selection of general horizon scanning as the decision-making process to be taken forward

In workshop 1 we also discussed potential outputs for the work. Two options were put forward:

a) A detailed blueprint for a hybrid decision support tool. The blueprint would be a realistic, actionable plan for a hybrid tool and service that addresses a specific decision-making process to enhance service quality and customer satisfaction. This would involve defining data sources, defining the roles, responsibilities and user actions, and mapping the processes and technology touchpoints.
b) A more speculative, high-level blueprint that incorporates 2-3 speculative future scenarios to inspire thinking in different directions. These scenarios would be a combination of narrative and visual design.The output would help to articulate and explore several potential future developments at a high-level.

Following discussions with our partners at Ofwat, we decided to move forward with a more detailed blueprint while also incorporating elements of visual narrative through a detailed scenario. Taking this approach was considered more useful as an input into their exploration of future opportunities for AI and/or hybrid intelligence.

Workshop 2 - designing the horizon scanning process for water regulators supported by HACID

In workshop 2, the goal was to look ahead and imagine the most effective and future-ready horizon scanning process by answering the challenge: **"How might a hybrid AI decision-making tool support water regulators and authorities in identifying emerging trends and risks?"** This workshop was intentionally speculative in nature. By taking participants into the year **2035**, the aim was to loosen current design constraints and spark creativity in addressing future challenges.

In Part 1, with a diverse group of water regulation experts from environmental, innovation, strategy, and policy teams, we began by crafting individual future statements, imagining how each participant, in their role, might benefit from using a tool like HACID. Through group voting, we selected the most relevant and promising decision in the context of our design challenge. This enabled us to move forward with the next steps: mapping key stakeholders and data sources. The final step was an attempt to outline a high-level process that could serve as a foundation for a future general horizon scanning framework.

Two groups developed distinct future scenarios for the horizon scanning process:

1. **How a Senior Associate in the Environmental Policy team** could use HACID to better guide water companies on where to focus their environmental improvement investments. The process has focused on a specific use case of tracing biological, ecological, chemical, and public health indicators to ensure surface water quality.

2. **How** a **Resilience Policy Analyst** could use HACID to be able to see a common operating picture of water and waste infrastructure and identify demand trends and gauge the preparedness against emerging hazards/threats.

These future-focused scenarios helped ground ideas in actionable contexts, laying the foundation for designing a robust decision-making process.

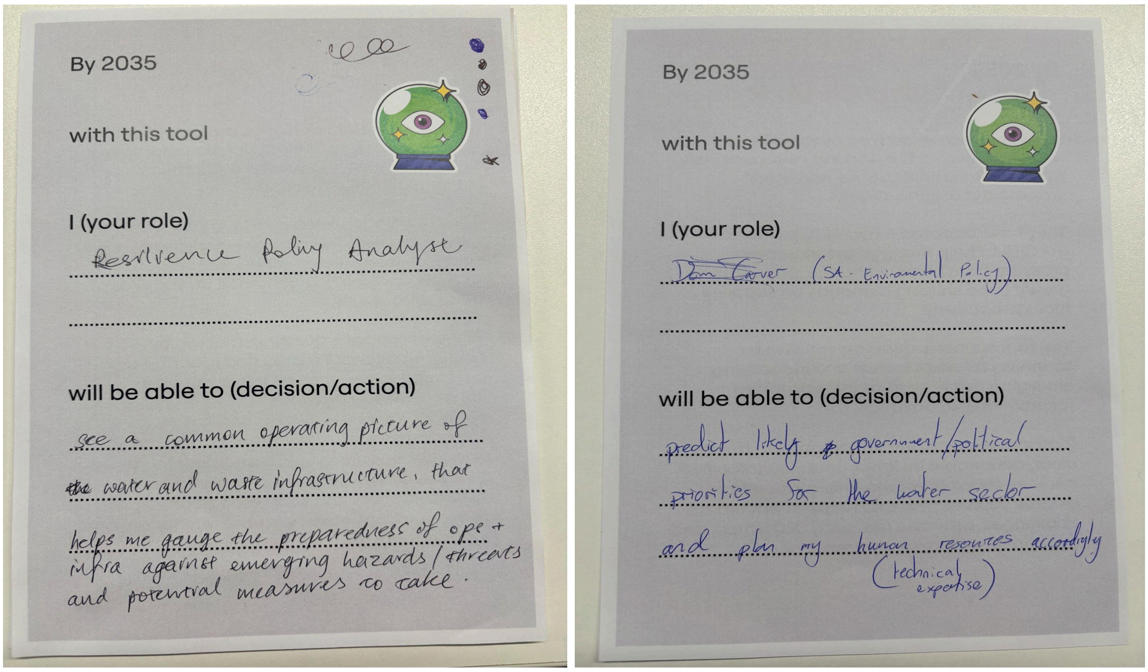


Image 3: These represent different decisions, aligned with two distinct water regulator roles, chosen by participants to be supported by the HACID tool.

What we learned:
The stakeholder landscape is complex, it also holds the potential for a truly cross-disciplinary decision making process and we identified a way to integrate this complexity into the blueprint implementing collective intelligence solutions. We successfully identified relevant data sources that could inform the future design and implementation of the HACID horizon scanning tool.

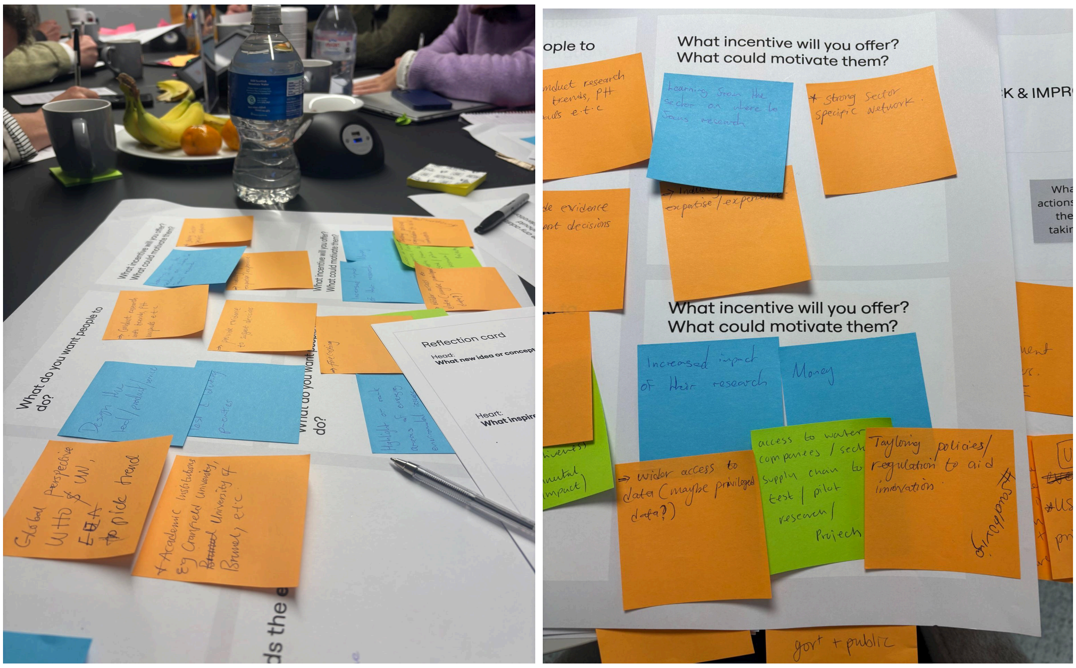


Image 4: This is an example of the "Mobile People" canvas, created in line with the methodologies outlined in the Collective Intelligence Playbook.

In Part 2 of the workshop, we focused solely on mapping what the horizon scanning processes could look like. We mapped out the key relationships and interaction points where the processes could deliver the most value, and concluded the workshop with two high-level maps of the two general horizon scanning use cases.

Image 5: This is one of two high-level draft blueprints outlining the decision-making process of a Senior Associate in the Environmental Policy team, aimed at better guiding water companies on where to focus their environmental improvement investments.

Blueprint overview and scenario description

The Blueprint serves as the first speculative, high-level design of the tool. While it is not yet detailed enough for technical implementation, it offers a foundation for future development. It illustrates the potential of how the tool could function by 2035 and provides a starting point for translating strategic vision into practical design. Initially we worked on developing two blueprints in parallel to reflect the two distinct pathways from part 2 of workshop 2. On further exploration of the viability of the blueprint related to the horizon scanning process for water and waste infrastructure and hazard preparedness we realised that the information about this use case was too limited to fully articulate a blueprint within the available time constraints. As a result, we created one final blueprint - see Figure 13 below.

The five phases of the horizon scanning process

Phase 1: Definition
In this initial phase, a representative from the water regulator uses the HACID tool to create a case and initiate the process of scanning demand trends in a specific location (e.g., South-East England). The aim is to build a clearer picture of future needs to support better-informed decisions on infrastructure planning, resilience, and investment.

Phase 2: Development & Data Collection
Eight representatives from key expert organisations review the submitted case and contribute relevant datasets and insights. These inputs help identify key factors that provide the essential context for the next step: scenario generation.

Phase 3: Foresight / Prediction
The HACID horizon scanning tool generates multiple future scenarios, enabling water regulators and other bodies to explore and debate different options for planning, resilience, and investment. These scenarios focus on long-term planning, with a 50-year outlook.

Phase 4: Visualise / Report Findings
A Resilience Analyst reviews the output from the HACID system, including an interactive map that visualises scenarios and recommendations over time, supporting clear communication of findings.

Phase 5: Feedback & Improve
Following scenario analysis and map exploration, future risks and opportunities become more evident. A cross-organisational strategy group is established to guide key decisions (e.g., price reviews), ensuring alignment with long-term trends and resilience goals.

How Water Regulators Might Use the HACID System

Built on an expert-driven knowledge graph co-created with sector specialists, the tool integrates critical environmental, societal, and economic datasets relevant to water management. This foundation enables HACID to identify emerging trends and risk - from net zero and biodiversity, to water demand, usage patterns, billing, and financial resilience - producing accessible, visually robust scenarios to proactively address the future challenges.

The tool facilitates the submission of diverse, open-ended cases, such as selecting optimal sites for future infrastructure, prioritising investments in catchment water quality, or guiding surface water quality monitoring. Water regulators are initially expected to submit approximately a dozen requests per quarter to support horizon scanning reports.
Specific use cases include:

- Assessing growing concerns around pharmaceuticals in waterways to inform the development of targeted policy responses.
- Exploring future challenges related to localised water demand and generating maps that illustrate the relative cost-benefit of investing in water quality improvements across catchments, thereby helping prioritise actions.
- Identifying which pollutants have the most significant impact in specific areas.
- Visualising the relationship between investments in storm overflow infrastructure and improvements in water quality, enabling more evidence-based decisions on future funding strategies.

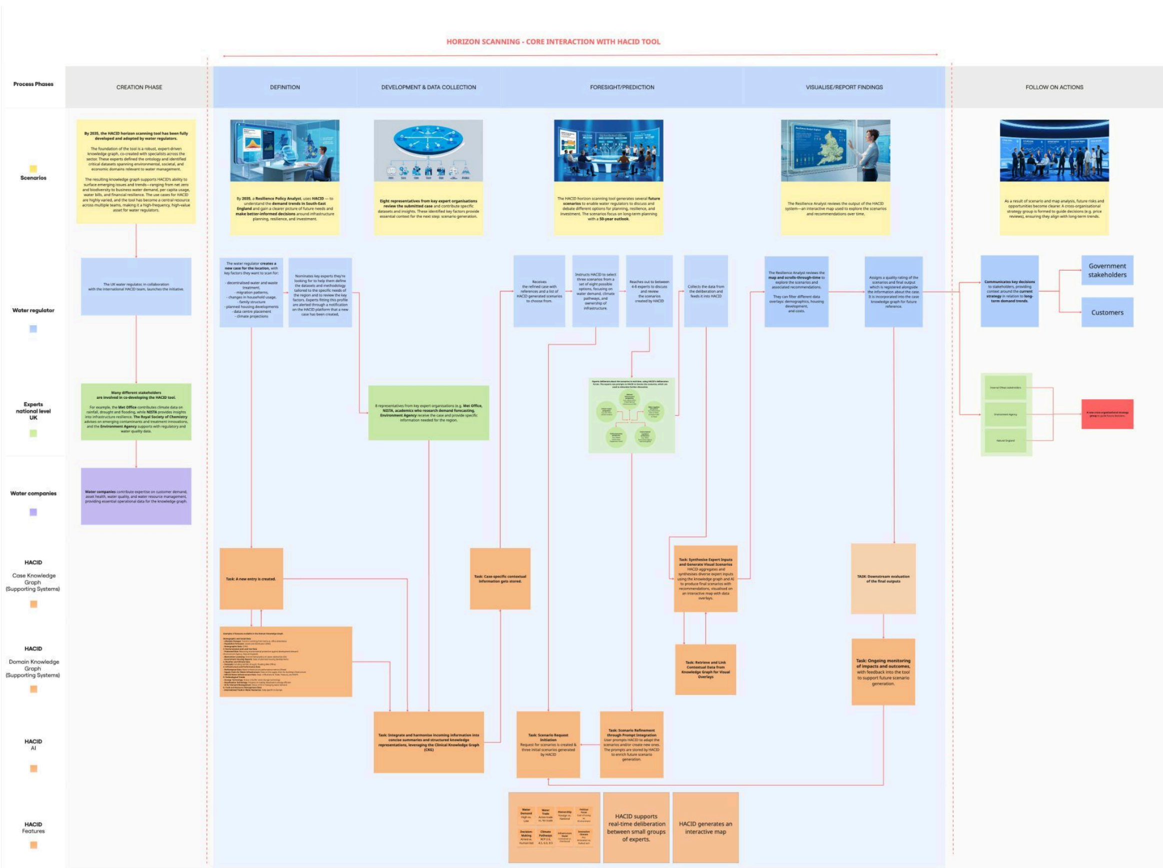


Figure 13: Blueprint overview

Conclusion on HACID for Water Regulation

Ofwat and other water regulators can use this blueprint as a foundation for strategic discussions around AI investments, helping them to identify priority areas where hybrid

human-AI systems like HACID could enhance decision-making, improve data integration, and better anticipate future risks in water management.
Potential challenges in building the HACID system include the complexity and fragmentation of the stakeholder landscape. The stakeholder environment in water regulation is described as highly complex, involving multiple organisations, data owners, and regulatory bodies with overlapping roles. This fragmentation presents several difficulties such as an incomplete and evolving data ecosystem. The effective use of HACID depends on the availability and quality of diverse, structured data.

3.3 Conclusion

The two alternative use cases in the energy and water regulation domains underscore the versatility and value of the HACID approach beyond its initial applications. They demonstrate how a hybrid system can effectively challenge traditional, top-down decision-making processes by integrating local, non-traditional data and empowering bottom-up decisions. By aggregating diverse inputs, from citizen-sourced reports to expert knowledge and quantitative data, HACID has the potential to enhance the effectiveness and transparency of decisions. However, three key challenges need to be addressed to ensure the successful implementation of this type of decision-support systems:

1. Viable incentive mechanisms to ensure high quality and sustained contributions by people;
2. The availability of human experts to construct and verify domain-specific knowledge graphs in the absence of established and widely accepted professional ontologies (which exist, for example, in the clinical domain); and,
3. The importance of verifiability and accuracy of information aggregated or generated through artificial intelligence or machine learning algorithms to ensure the systems are trusted by domain experts.

Future research and innovation programmes that foreground hybrid intelligence approaches that combine crowdsourcing, artificial intelligence (using both classic machine learning and LLM-based approaches) and knowledge graph technology should focus on addressing these challenges.

4. Appendix

4.1 Acknowledgements

We are grateful to all individuals who contributed their expertise during our research. In particular, we would like to thank the GSA MDes students and their tutors for their work and contributions to the research and design assets in Section 3.1: Gabriella Moretti-Miles, Sophie Anne Val, Leo Malins, Oona O'Brien, Noah Albrectsen, Kirsty Ross, Dr Marianne McAra, Zoe Prosser and Studio AndThen. We are grateful to the following collaborators at partner organisations who helped to convene stakeholders and to deliver roundtable discussions: Seth Bullock (AI4CI Hub), Lara Groves (Ada Lovelace Institute), Fai Fung and Neha Mittal (Met Office).

4.2 Overview of activities and participating organisations

Policy Roundtable	Format, Date	Participating organisations
Practitioner Perspectives: insights into participatory AI In collaboration with Ada Lovelace Institute	29 April 2024, In-person workshop, 2.5 hours (London)	Microsoft Research; Cohere for AI; Incubator for AI (UK Government); Westminster City Council; IBM Watson; Department of Work & Pensions, UK Government; Cabinet Office, UK Government.
AI Regulation for Clinical Decision Support – From Principles to Implementation In collaboration with the AI4CI Hub.	28 April 2025, Online workshop, 1.5 hrs	MHRA; AI4CI Hub; NHS England; University of Bristol; Alan Turing Institute; University of Ulster
AI Governance and Trust in Climate Services	21 July 2025, Online workshop, 1.5 hrs	Met Office representatives: International team: Applied sciences and services Climate Services team Product team Science fellows x2 Programme Delivery team

Alternative HACID use case - Energy systems

Activity	Date & format	Participating organisations
Interviews	Early October 2024	ISTC-CNR x1, Nesta: Sustainable Futures team x3; Other: anonymous energy experts (Scotland).
Co-creation workshop	30th October 2024	ISTC-CNR x1, Nesta: Sustainable Futures team x3, Nesta: Centre for Collective Intelligence x2
Showcase	26th November 2024	ISTC-CNR, Met Office, Max Planck Institute, Nesta: Sustainable Futures team, Nesta: Centre for Collective Intelligence, Glasgow School of Art

Alternative HACID use case - Water regulation and management

Activity	Date & format	Participating teams
Scoping meeting	4th March 2025. Online discussion. 0.5 hrs.	Policy & Environment team. Innovation team.
Workshop 1	24th March 2025. Online session. 1 hr.	Policy & Environment team. Innovation team.
Workshop 2	15th April 2025. In-person workshop. 4 hrs.	Policy & Environment team. Innovation team. Operational resilience team. Government engagement team.