Essays on AI and professional practice

Will Allen

Working well in the age of AI

Judgement and responsibility in professional and collaborative practice

learning for sustainability
supporting engagement, co-design and adaptation

learningforsustainability.net

Working well in the age of AI
Judgement and responsibility in professional and collaborative practice

Will Allen
September 2026
Ōtautahi Christchurch, New Zealand
willallen.nz@gmail.com
learningforsustainability.net

Cover and document design: Sanja Stojanović
Figures and graphics: Will Allen, with ChatGPT assistance

This collection brings together essays on the use of generative AI in professional and collaborative practice. It focuses on judgement, participation, interpretation, learning and responsibility, and on what needs to remain visible and open to discussion as AI becomes part of everyday work.

The essays draw on material developed through the Learning for Sustainability website. They are intended to support reflection and discussion rather than provide a fixed framework for AI use.

Version 1.0
DOI: 10.5281/zenodo.22700603

Suggested citation: Allen, W. (2026). *Working well in the age of AI: Judgement and responsibility in professional and collaborative practice*. Learning for Sustainability. https://doi.org/10.5281/zenodo.22700603

Contents

Preface

Why engage with AI at all

There are good reasons to be cautious about the development and use of generative AI. Concerns include its environmental and social costs, the concentration of power in a small number of technology companies, the use of data and creative work in developing AI systems, effects on employment and professional practice, and the wider consequences of increasingly capable systems. For some people, these concerns lead reasonably to the view that some uses of AI should be resisted or avoided altogether.

This collection starts from a practical, but compatible, position. Generative AI is already becoming part of professional and organisational life. That makes it important to ask not only whether it should be used, but what happens when it is used: what it changes about how we think and work together, whose knowledge and perspectives remain visible, where judgement sits, and who remains responsible for the consequences.

These essays grew largely from my own efforts to understand those questions through practice. They were written over roughly 18 months, beginning in early 2025, during a period in which both the technology and my own use of it were changing quickly. I have revised them for this collection but have retained something of that development rather than rewriting them as though they were all produced at the same point in time. They are not an argument for adopting AI, nor a comprehensive framework for its responsible use. Rather, they explore what people, teams and organisations might need to pay attention to as AI becomes part of research, evaluation, facilitation, planning and other forms of collaborative work.

A concern running through the collection is what can become less visible around apparently capable outputs. Judgement, participation, interpretation, learning and responsibility do not disappear when AI is used, but they can become easier to overlook. The essays ask how these aspects of professional and collaborative work can remain visible, discussable and open to judgement.

Although the collection is grounded in professional practice, it does not stop there. Some essays consider how teams and organisations might develop shared expectations, build capability, guide AI use and learn from experience. Others step further back to consider the organisations developing and deploying AI, and the wider economic, political and ecological systems within which they operate. The later essays ask what better AI design and development might require, including attention to purpose, human agency, fairness, planetary limits, wider system effects and accountability.

The practical starting point is deliberately modest: if AI is becoming part of our work, how might we work with it without losing sight of judgement, participation, learning and responsibility? Sometimes the answer may be to use it. Sometimes it may be to use it differently, constrain its role, or decide that it should not be used at all. But these choices are also shaped by organisational conditions and by how AI systems themselves are designed, developed and governed.

Introduction

Working well across practice, organisations and wider systems

Much discussion of generative AI concentrates either on what the technology can do and how it can be used, or on its wider ethical and societal implications. Both matter. These essays are particularly concerned with the space between them: what happens when AI becomes part of professional and collaborative practice, and how everyday use connects with organisational choices and wider systems.

In professional and collaborative work, some of the important questions concern what happens around the output. Who interprets it? Whose perspectives are included or overlooked? What happens to discussion, judgement and learning? Who remains responsible when AI-generated material influences a decision, a strategy, a meeting, an evaluation or a piece of research?

A recurring concern in these essays is that AI can produce increasingly convincing outputs while making some of the work needed to judge them less visible. Inquiry, participation, interpretation and learning may be reduced or displaced even while the resulting artefact appears more complete. The challenge is therefore not simply to keep a human involved, but to retain the experience, context and opportunity for judgement needed to recognise when an output is good enough to rely on.

This matters particularly in interpretive and collaborative work, where the quality of an outcome depends not only on the artefact produced, but on the inquiry, participation and shared judgement through which it was developed.

These questions become more important as AI moves from being an occasional tool to becoming part of everyday work. The issue is not simply whether a system gives a useful answer. It is how its use changes the way people frame problems, make sense of information, work with others and decide what to do next.

This collection looks at those questions across three connected levels.

Figure 1: Three connected levels for considering ethical concerns around AI development and use. Practice, organisational and structural concerns overlap, but each brings different questions into view.

The distinctions are useful because different ethical concerns, forms of responsibility and kinds of judgement come into view at each level, even though decisions and effects often move between them. Ethics is therefore not a separate layer added to questions of AI practice or governance, but something that runs through all three levels.

At the **practice level**, the focus is on how individuals and groups use AI in day-to-day work. This includes questions of judgement, interpretation, authorship, participation, learning and responsibility.

At the **organisational level**, the concern shifts to how teams and institutions set expectations, develop guidance, build capability and decide what kinds of AI use are appropriate in different settings.

At the **structural level**, wider questions come into view: who designs, develops and controls AI systems, whose interests they serve, how benefits and burdens are distributed, and what environmental, social and political effects accompany their development and use.

These levels overlap, but they are not interchangeable. Better practice cannot resolve every structural concern. Organisational policy cannot on its own address the wider political economy of AI. Equally, structural critique does not remove the need for people and organisations to make practical choices about systems that are already entering their work.

The collection is grounded in practice, but moves deliberately across all three levels. Part 1 looks at what better AI-supported work might mean and how AI is already entering professional practice. Part 2 turns to shared and collaborative work, where questions of authorship, participation, trust and collective judgement become more visible. Part 3 considers learning, capability and practical organisational responses. Part 4 then steps back to the wider ethical and structural picture and asks what better AI design, development and use might require.

This movement from practice towards organisations and wider systems is deliberate. Practice is where many of the immediate consequences of AI use become visible, but it is not the boundary of the collection. Grounded experience can also help organisations develop better guidance and governance, while wider structural concerns raise questions about how AI itself is designed, developed and governed.

Where useful, the essays also link to related material on the Learning for Sustainability website, allowing readers to follow particular ideas into the wider body of practice resources from which the collection has developed.

The essays can be read from beginning to end, but they do not need to be. Each stands on its own and can be approached according to the reader's particular interests. Together, however, they return to a common concern:

> **What needs to remain visible, discussable and open to judgement when AI becomes part of professional and collaborative work?**

That question runs through the collection more consistently than any single framework or set of recommendations. The answers will differ across contexts.

PART 1
Starting with practice and judgement

AI is already becoming part of everyday professional work, often through small experiments with drafting, research, analysis and preparation. This section begins with that practical reality. It asks what better AI-supported work might look like and then reflects on AI as a thought partner in work that depends on judgement, interpretation and systems thinking. The focus is not simply on what AI can produce, but on how it can support good work without weakening the thinking, responsibility and professional judgement on which that work depends.

1. What would better work with AI look like?

Much discussion about AI focuses on models, governance and regulation. This essay looks instead at what happens to professional practice when AI becomes part of everyday work. It considers how practitioners, teams and smaller organisations can decide what better work would mean. It explores how they can keep useful outputs connected with the inquiry, participation and judgement that make the work worth trusting, and learn where AI use should continue, change or stop.

AI use rarely begins with a strategy.

More often, someone tries a tool to help with a report, a meeting summary, an early analysis or a difficult first draft. A colleague sees the result and experiments in turn. Before long, prompts are being exchanged, parts of workflows are changing and expectations are beginning to shift. Some uses are discussed; others remain private.

By the time an organisation develops an AI policy or strategy, generative AI is often already influencing how work is framed, produced, reviewed and valued. Repeated individual choices become team habits, informal norms and management expectations. Work that once took a day begins to be expected in an afternoon. A polished first version becomes the accepted starting point.

The issue is not only that these uses can be difficult to see. AI can produce a recognisably useful report, synthesis or analysis while loosening its connection with the process that gives it value. The output can look complete even though the inquiry behind it has narrowed, important differences have not been worked through, or interpretive choices have become harder to examine.

In many settings, the practical question is what happens to professional practice when AI is already part of it. How can uses already under way contribute to better work without quietly redefining what counts as enough?

A useful first step is to make existing practice visible: where AI is being used, what it is changing, and where useful outputs risk becoming detached from the work that gives them value.

Clarify what better work would mean

AI discussions often begin with what a tool can do. A better starting point is the work itself. What are we trying to improve? For whom? What is not working well now? What would people notice if the work became better?

Better work is rarely defined by speed alone. Depending on the setting, it may mean better judgement, clearer communication, stronger analysis, more thoughtful decisions or more time for relationships, interpretation and learning. For a small organisation, it can also mean doing work that existing resources would otherwise place beyond reach.

The intended contribution differs across settings: more time to examine competing explanations in research; a clearer account of where participants agree and differ in facilitation; or faster handling of routine enquiries without losing sight of unusual or higher-risk cases in a service. These are more meaningful reference points than uptake or output volume.

AI can strengthen one part of the work while weakening another. A faster synthesis may leave less time for questioning assumptions. A clearer summary may smooth over disagreement. Consistency may be useful, but it can also discourage professional judgement.

AI may also redistribute work rather than simply reduce it. Time saved in one activity may reappear later through checking evidence, questioning interpretation, coordinating decisions or explaining how conclusions were reached. Looking only at the speed of one task can therefore give a misleading picture of whether the wider work has improved.

The question is whether the wider work becomes better.

“Better” will also be contested. Faster work may be less reflective. Wider access may introduce new errors. A more consistent process may be easier to manage but less responsive to context. The aim is not to settle on a universal definition, but to keep the intended contribution, choices and trade-offs open to discussion and review.

Notice what AI is changing around the work

AI can strengthen a process. It can help people surface overlooked material, make assumptions visible enough to question, or explore alternatives that a small team could not otherwise afford. An initial synthesis can become a useful object for collective scrutiny rather than a conclusion to accept.

The same output can also change the surrounding practice in less helpful ways.

One change concerns framing. AI can quickly produce a first account of a problem, a set of themes or an apparent synthesis. Because the result is fluent and well organised, it can become the basis for what follows. Later contributors respond to its categories rather than revisit how those categories were formed. As discussed in Essay 4, *AI and the authority of the first synthesis*, I have seen how quickly a clear early synthesis can become the document everyone works from, even when its categories were meant to be provisional.

Another change concerns participation. Essay 5, *AI can produce the record of participation without the participation*, considers how AI can create summaries, personas, plans and coherent accounts of what people need. The visible artefacts can be present even though the relationships, disagreement and shared learning through which understanding normally develops have not taken place. The question here is how teams can notice that separation and reconnect the output with the people and processes it represents.

Judgement can also become harder to see. Someone may still approve the final output, but important interpretive choices have already been made through the prompt, the material supplied, the framing produced and the selection of what to retain. Human review then risks becoming a check of the polished result rather than an examination of how it came to be. As Essay 4 explores more fully, the harder question is whether the process still allows people to question the framing and exercise judgement together.

Workload adds another complication. Time saved does not automatically become time for reflection, engagement or higher-value work. It can simply reset expectations. A team that can produce five reports in the time previously needed for three may soon find that five are expected.

Across these examples, AI provides a recognisably useful result while loosening its connection with the process that gives it value. The synthesis can be ready before its framing has been tested; a complete-looking record can stand in for differences that were never worked through; time saved can disappear into the next deadline. The issue is whether the output remains connected to the inquiry, interpretation, discussion and testing needed to use it well.

A useful output is only one part of better work. The diagram below shows some of the connections that need to remain visible around it.

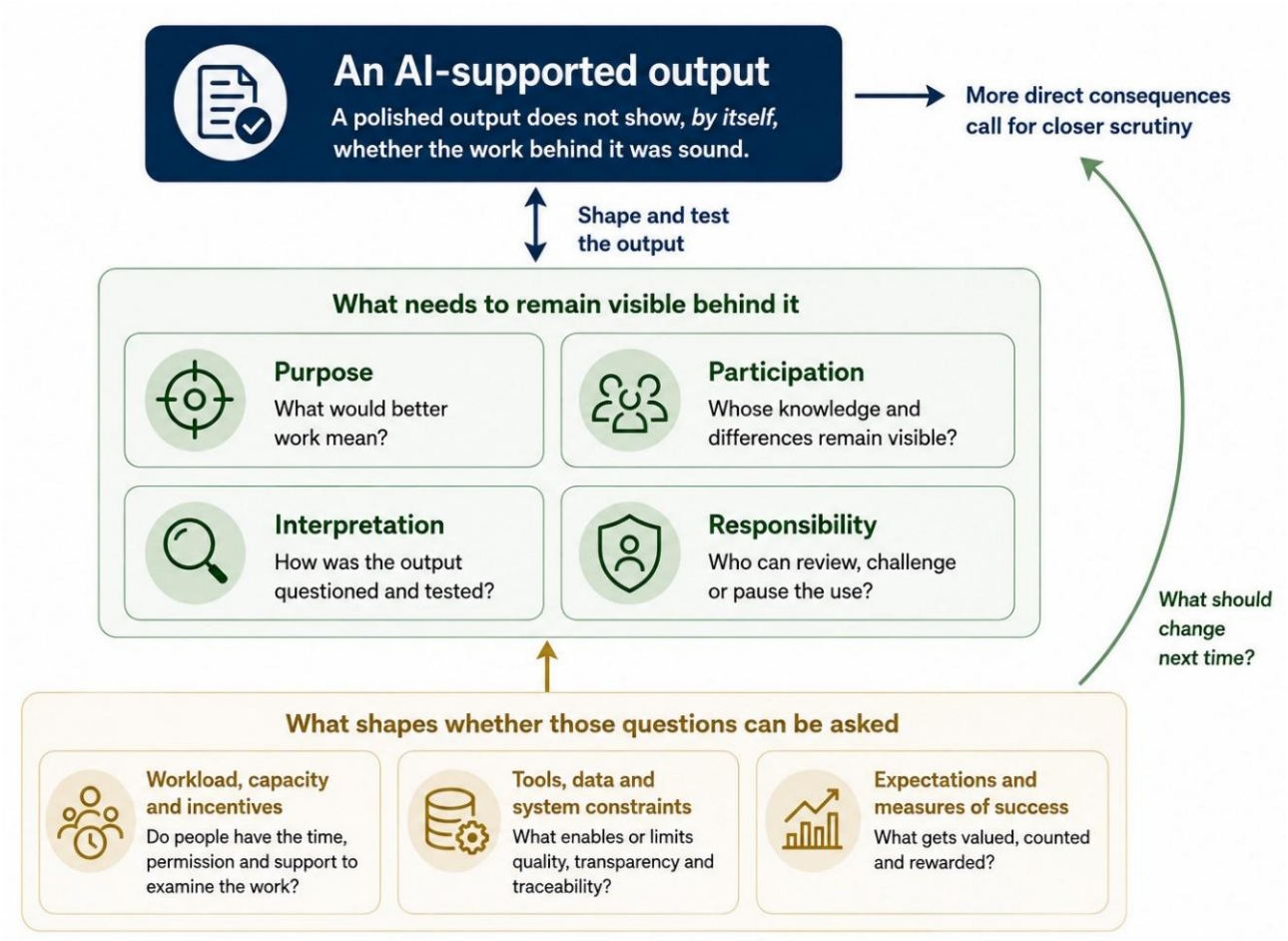

Figure 2: Looking beyond the output. A polished AI-supported result does not, by itself, show whether the wider work was sound.

The diagram distinguishes four parts of the work that need to remain visible: purpose, participation, interpretation and responsibility. It also shows how workload, incentives, tools and organisational expectations shape whether people have the space and permission to question what is being produced. These elements interact throughout the work rather than forming a fixed sequence.

This wider view has parallels in sociotechnical approaches to AI evaluation. Weidinger and colleagues (2023), for example, distinguish between evaluating system capabilities, human interaction with AI, and wider systemic effects.

Pay closer attention when effects travel further

These questions become more demanding, and shortcomings harder to correct, when AI use reaches beyond an internal draft and becomes part of a product or service.

Consider a service team using AI to classify incoming enquiries. Routine cases are sorted more quickly and response times improve. But better service also depends on unusual cases remaining visible, staff being able to question the classification, and people having a clear route to human support.

If those conditions do not hold, the system can perform well against average response time while making the service worse for people whose circumstances do not fit common patterns.

When AI shapes how a service responds or decides, trust depends on reliability, transparency and clear responsibility. Uses that directly affect people therefore need closer scrutiny, clear routes for review, and a practical way to pause or withdraw them.

Build the capability and conditions to question practice

Organisational discussions about AI capability often focus on access, literacy and prompting skills. These matter, but they are not enough. People also need to recognise hidden uncertainty, trace important claims, question an output's framing and explain how AI influenced the work. These concerns sit alongside privacy, confidentiality, security and factual reliability. They also draw attention to aspects of good work that technical and compliance checks often miss.

Capability develops through real work. Teams can examine actual examples, including weak or ambiguous ones, compare AI-assisted outputs with the underlying material and discuss what was lost, added or left unresolved. These arrangements need not be elaborate: selected case reviews, clear expectations about disclosure and additional scrutiny where errors could materially affect the work may provide a useful starting point.

The harder issue is whether people have permission to question the practice. If every discussion of AI is tied to productivity gains, staff are less likely to surface errors or admit that an experiment failed. If disclosure feels like surveillance, use remains private and unexamined. If managers reward speed and volume, careful review becomes harder even when guidance says it matters.

Some conditions are set beyond the immediate team. Tool design, commercial incentives, organisational targets, employment conditions and regulation all shape what responsible use is possible. Local practice cannot resolve all of these pressures, but it can make them visible. That helps avoid treating every problem as a matter of individual skill.

Responsibility also needs practical expression. Who decides that AI is suitable for a task? Who reviews uses carrying greater risk? Who can pause them? Who remains accountable for the interpretation?

In participatory or interpretive work, participants, facilitators and others close to the process may be better placed than those producing the output to notice that an important difference has been flattened, tentative agreement has become a firm conclusion, or the language no longer reflects what people meant.

The point is narrower than involving everyone in every decision. Judging quality sometimes requires knowledge held by the people represented in, affected by or expected to act on the work.

Learn together and adjust the practice

Experimenting with AI becomes organisational learning only when experience can be shared, compared and used to change what people do.

In one setting, a practitioner may find that AI helps generate questions during early exploration but is less useful for synthesising contested evidence. Elsewhere, a team may find that it improves a report's structure while requiring more checking than expected. These are lessons about particular uses, materials and working arrangements, not fixed conclusions about what AI can or cannot do.

If those lessons remain with individuals, other teams may repeat the same experiments and encounter similar difficulties.

Shared review should go beyond asking whether a tool worked. Essay 6, *How do we know whether AI is improving the work?* looks more directly at how evaluation needs to follow what happens around and after an output, not only whether the output itself was satisfactory.

Three questions are enough to begin:

- What became better, and for whom?
- What changed around the work that we did not anticipate?
- What should we continue, adapt, limit or stop?

These questions reconnect the intended contribution with evidence about what happened and decisions about what should change. Teams do not need a large dashboard for every use, but they do need enough evidence and shared judgement to distinguish a useful experiment from a practice that merely produces convincing outputs.

These are familiar disciplines within planning, facilitation, evaluation and organisational learning. AI does not replace them, but useful-looking outputs can make them easier to overlook.

A team may find that AI helps organise a large body of material before analysis, but is less suitable for preparing the account shared with participants or decision-makers. It may retain the earlier use, change how the output is reviewed, or involve others where interpretation remains uncertain or contested.

That is not necessarily a failed experiment. It is a more precise understanding of where the use contributes and where it does not. Learning may support wider use, but it may also lead to a narrower role, additional review, use only under particular conditions, or stopping altogether.

Working well with AI

Practitioners and smaller organisations can act more deliberately without waiting for a complete strategy. Essay 7, *Working with AI: where to begin?*, returns to this question from an organisational perspective. AI use is already developing through ordinary choices about drafting, research, analysis, communication and decision-making. Those choices gradually establish what an organisation will

accept as sufficient process behind an output. How much checking, interpretation or involvement is expected is rarely decided explicitly. More often, it is settled through practice.

The important part is to keep those choices open to question and to learn from experience before early trials quietly become established practice. The aim is to keep useful outputs connected to the work that makes them worth trusting, rather than treating wider adoption as the measure of success.

AI will continue to evolve, and it can help people work more quickly and extend what small teams can do. The professional disciplines that help people work well together change more slowly: being clear about purpose, examining assumptions, preserving different perspectives, evaluating consequences and learning from experience. Better work requires more than useful results. It requires processes people can examine, learn from and stand behind.

2. AI as a thought partner: reflections on collaborative practice and systems work

This essay shares practical reflections on how AI tools can support those working in contexts where collaboration, co-design, and facilitation are important. Drawing from experience, I outline four ways these tools have become quiet thought partners in my work – and reflect on where their strengths end and human judgement, relationships, and systems thinking still matter most.

Since early 2023, I have used tools such as GPT, Claude and Perplexity across a wide range of projects, mainly to support framing, exploration, writing and preparation rather than to outsource my thinking. In practice, I've found AI tools to be especially helpful when working through ideas – whether I'm drafting a blog post, curating material for the Learning for Sustainability site, designing a workshop, or synthesising interviews and field notes. They help me surface insights, test interpretations, and gain clarity more quickly.

For those of us working in participatory, systems-facing settings, much of the work is relational. It involves framing, convening, and supporting collaborative thinking and action. AI doesn't replace that, but it can help us prepare more thoughtfully for it.

Looking across my own practice, four uses have become particularly helpful.

Exploring new research and perspectives

As someone who curates open-access resources for the Learning for Sustainability site, I'm regularly reading across disciplines. But when I need to explore a new topic, trace emerging framings, or look for fresh insights, GPT and Perplexity have become helpful companions.

With the right prompts, they can surface recent academic articles, policy papers, and blogs that might not turn up in a standard search. They can also suggest related concepts, authors or bodies of work that I may not have thought to look for, which can help me move beyond my initial framing of a topic. These tools can still invent, misidentify or misrepresent sources, so I check and read the sources before drawing on them.

It's not a shortcut for reading. The value is more in broadening the field, helping me see where else to look, and sometimes showing me that the question I started with was too narrow.

Supporting systems framing and complexity

Framing, not fixing. In systems work, I often find myself dealing with issues that don't have tidy answers. Whether I'm preparing a Theory of Change, shaping a facilitation run-sheet, synthesising interviews and field notes, or developing reflective prompts for group work – these are all acts of framing, part of the wider systems thinking and systemic design practice I describe here.

AI can surface possible framings, contrast assumptions and reveal tensions I might otherwise miss. I still need to decide what is useful, what fits the context, and what I am prepared to stand behind. Different framings can make different relationships, trade-offs or blind spots more visible.

Sometimes I use it to trial different starting points for a session, or to test how a set of themes might cluster. Other times it helps me rehearse arguments or explore ways to present a tricky issue. The back-and-forth can help me see where a framing is too narrow, where an assumption needs questioning, or where another perspective might change the way the issue is understood.

A recent example came while developing this collection. I had used AI repeatedly to review the essays and test how well they worked together. The AI reviews tended to see the collection mainly through its concerns with professional practice, judgement and participation. What they did not surface was the stronger pattern that had developed across the collection: a movement from practice, through organisational conditions, to wider structural questions. Once I recognised and named that framing, AI became useful again in testing and strengthening it. The experience was a useful reminder that a thought partner can help explore a frame without necessarily recognising the frame that matters most. Part of my role as a writer is still to notice the larger pattern, decide what is significant, and sometimes reject plausible advice because it does not serve the whole.

Making writing less lonely

When writing, I often start with scattered notes and phrases. I may use AI to explore possible structures or ways of expressing an idea, reject much of what comes back, follow a connection I had not thought of, and sometimes rethink the argument altogether. In that kind of exchange, it is not always possible, or particularly useful, to draw a clean line between where one contribution ends and another begins. Once the direction becomes clearer, I work closely through the structure, paragraphs and sentences, although that can still send me back to reconsider the thinking.

For me, the value lies partly in that back-and-forth. The important question is whether I am still questioning, making connections, changing my mind and working out what I actually think. By the time I publish, I need to be satisfied that the piece says what I mean, that I have worked through the argument rather than simply accepted it, and that I am prepared to take responsibility for it.

This is not entirely different from writing with co-authors, a question I return to in Essay 3, *Working with AI in the room: authorship, responsibility, and collective judgement*. In a good collaborative process, ideas develop through discussion, challenge, rewriting and response. After several rounds, it can be difficult to say exactly whose idea a particular point was, or how it evolved. What matters is that the thinking has been actively worked through, rather than simply accepted because someone, or something, proposed it.

This way of thinking about AI as part of a collaborative process is also beginning to appear in research on human–AI learning and writing. Samuel (2026), for example, uses the idea of *epistemic co-agency* to describe forms of human–AI interaction in which people reason with AI while retaining responsibility for knowledge construction. That is close to the role I am describing here: AI can contribute to the development of ideas, but judgement about meaning, relevance and what to stand behind remains human.

Testing how things land

Another way I use these tools is to explore how a piece of writing might be interpreted by different audiences. I might ask: *How might this come across to a biophysical scientist? A social researcher? A policymaker? A practitioner working on the ground?* I would not take these responses as reliable predictions of how actual people will respond, but rather as one way of stretching perspective and looking for possible blind spots.

This has become a regular part of how I write and review, especially when preparing public-facing material or strategic documents. It can help me clarify tone, consider possible misreadings, and think about whether the language could be more inclusive. But it can't stand in for lived experience or cultural ways of knowing, and it's important to stay mindful of that, especially when working in equity-focused or cross-cultural spaces. This kind of reflective adaptation is key to how I think about co-design in complex settings.

Working with AI tools, I've found it important to think carefully about the ethical implications, especially around data sensitivity, privacy, and being transparent about when and how these tools are used. I'm also mindful of concerns around AI-generated content and source attribution. This includes being alert to the risk of unknowingly drawing on material that hasn't been shared with permission. In work that values transparency and shared knowledge, these are practical ethical questions that need to remain in view.

Working in complementary ways

The examples above highlight how AI tools can support real work without replacing it. In practice, I've found it helpful to think about where these tools complement – rather than compete with – human judgement, facilitation, and collaboration. The table below sets out some of those distinctions, based on how I use AI in day-to-day practice.

Table: How AI can contribute to collaborative research and facilitation

Task/Role	How AI might contribute	What still needs human judgement
Literature search & synthesis	Rapid scanning, surfacing new sources	Critical reading, contextual judgement
Qualitative analysis & synthesis	Summarising interviews, clustering themes	Meaning-making, context, ethical interpretation
Framing & systems thinking	Suggesting framings, testing assumptions	Coherence, relevance, group-based interpretation
Drafting & rewriting	Contributing potential structures, arguments, connections and wording	Questioning, selection, revision, voice, meaning and responsibility
Testing audience response	Simulating reactions, checking tone	Knowing what matters, inclusive framing, cultural nuance
Facilitation & relationship work	Contributing prompts, synthesis and other inputs within deliberately designed group processes	Trust, empathy, power dynamics, situational judgement

What remains ours

Looking across that table, AI can contribute to analysis, drafting and exploration, but it does not take on the human responsibilities that sit at the heart of participatory and systems-facing work.

AI does not itself hold relationships, understand what is at stake for people, or take responsibility for navigating power and consequence. Those remain human responsibilities. They require trust, timing, emotional intelligence, and a deep understanding of context.

In collaborative research, evaluation and systems change work, generating insight is only part of the task. We also need to notice patterns and tensions, help people make sense of complex situations, and support decisions that those involved can understand and work with.

For me, these technologies are most useful when they support the thinking, judgement, relationships and learning that the work depends on. That becomes more complicated when AI moves from individual use into work that others are expected to interpret, trust and stand behind.

PART 2
AI in shared and collaborative work

When AI enters shared work, its effects extend beyond individual use. It can shape authorship, participation and trust, influence how issues are framed, and affect what a group comes to treat as settled. This section looks at those changes, and at how collaborative processes can keep interpretation, disagreement and collective judgement open to the people involved.

3. Working with AI in the room: authorship, responsibility, and collective judgement

As AI moves from individual use into meetings, workshops, strategy sessions, and collaborative writing, it increasingly enters the room with us. Some find this useful, while others experience unease, often voiced in terms of the tools themselves. Looked at more closely, that unease tends to centre on authorship, responsibility, trust, and how collective judgement remains legitimate when AI-generated material becomes part of shared work. This essay focuses on those practice-level tensions and what they mean for working together well.

Essay 2, *AI as a thought partner: reflections on collaborative practice and systems work*, considered how AI can support individual reflection and preparation without replacing judgement. This essay moves the focus outward, to what happens when AI becomes part of shared work.

AI may also enter shared work through participants themselves, for example in preparing, translating or developing a contribution. This may support participation, while also changing what a contribution can reasonably be taken to represent.

These questions are not easily resolved through formal policy or agreed conventions alone. They signal that familiar practices of judgement and shared sense-making are being stretched. Paying attention to where discomfort shows up can help teams and facilitators work with AI in ways that remain supportive, rather than quietly authoritative.

These tensions do not all show up in the same way. They tend to become visible around authorship, responsibility, trust, and the effects of convenience on engagement. They overlap, and together shape whether people can still recognise, question and stand behind the judgement that emerges from shared work.

When help becomes authorship

One of the first places unease shows up is around authorship, particularly as AI moves from individual use into shared work. When people work with AI on their own, the boundaries between their own contribution and the model's can already become difficult to draw. In collaborative settings, that question becomes more consequential because authorship and judgement are shared among people as well.

One uncertainty is whether work shaped with AI still feels like something people can stand behind. This is especially visible in writing-heavy roles, where contribution is closely tied to voice, craft, and judgement. When AI is used to restructure or polish shared documents, some people experience a subtle loss of recognition for their contribution, even when the content itself remains sound.

In shared work, authorship is less a label than something negotiated through the process. In practice, however, it often is not. Roles, hierarchies, and established conventions frequently stand in for open

discussion of contribution. Introducing AI can intensify existing tensions, and sometimes create new ones, particularly where people are already uneasy about its role in shared work.

Rather than trying to draw hard lines around what "counts" as human or AI-assisted work, teams may be better served by slowing down moments where authorship matters most. This includes points where work is shared externally, where commitments are being made, or where people are expected to stand behind particular claims. Naming how a piece of work came together, including the role of tools, can help restore a sense of agency and mutual recognition without turning transparency into surveillance.

Responsibility without ownership

A second tension shows up around responsibility. AI can generate options, summaries, and recommendations quickly, but cannot hold responsibility for them. In principle, accountability remains human. In practice, that clarity often feels thin.

A similar concern appears in Green's (2022) critique of human oversight of algorithmic decisions. Although his focus is government use of algorithms, he argues that retaining a human role does not by itself ensure meaningful oversight if people cannot effectively exercise the judgement expected of them.

> **A useful test is whether people still feel able to say, with confidence, "this reflects my judgement" or "this reflects our shared thinking."**

A more difficult issue is the gap between formal responsibility and felt ownership of decisions. AI makes it easy to surface plausible answers faster than individuals or teams can fully work through the judgement involved, particularly under time pressure. In strategy and planning contexts, AI-generated frameworks or draft plans often enter the room already appearing complete, leaving limited space for shared sense-making. While often intended as preparation, this can feel to others like the difficult work of thinking together has already been done elsewhere.

Over time, this creates unease. People are effectively asked to endorse outcomes they did not meaningfully shape, or to stand behind recommendations that feel imported rather than worked through. When things go wrong, accountability snaps back sharply onto individuals, reinforcing the sense that responsibility has been shifted without being shared. The distinction I find useful here is between responsibility and participation. Responsibility in collective work is not only about being answerable after the fact, but about having had a hand in shaping the judgement that led to an outcome.

This work does not take place in a vacuum. Time pressure, organisational incentives, accountability regimes, and performance expectations often limit opportunities to slow down, reopen judgement, or create space for shared sense-making, even when people recognise the need to do so. Within those constraints, teams may need to pay particular attention to moments where judgement is re-entered. Slowing down adoption, reopening assumptions, or asking what has not been considered can help restore agency before decisions harden. When people can say, "I understand how we arrived here," responsibility feels shared and credible rather than procedural.

Trust in shared spaces

Trust is often tested when AI outputs enter shared spaces such as meetings, workshops, and facilitated sessions. Questions also arise about whether and how to disclose AI use in these settings. Bringing AI-generated summaries or slides into a meeting can feel risky if errors or misinterpretations are exposed. At the same time, not naming AI's role can feel deceptive, particularly in contexts that depend on openness and credibility.

This tension becomes sharp around records of collective work. AI-summarised meetings, action lists, or reflections often appear tidy and authoritative. Yet a tidy record can still leave participants feeling misrepresented. Nuance, disagreement, and relational dynamics are easily flattened. When that happens, it becomes harder to contest the record or revisit decisions, especially if the summary is treated as neutral or final.

For facilitators and process-holders, this raises questions of legitimacy. Shared spaces rely on participants trusting that their contributions are heard, represented with care, and open to challenge. When AI quietly shapes what is captured and remembered, it can shift narrative power in ways that are difficult to notice.

I would therefore treat shared records as provisional and open to revision. Long before AI, notes and summaries shaped what groups remembered and acted on, sometimes unevenly and with contested effects. AI does not change that dynamic, but it amplifies it by producing outputs that sound confident and complete.

Accuracy matters, but so does contestability: can people still question, correct and revise the record? Making it explicit that summaries are starting points, inviting correction, and keeping ownership of interpretation clearly human helps protect the integrity of shared spaces. Trust is sustained not by perfect capture, but by the ongoing ability to question and revise what has been recorded.

Erosion through convenience

A final tension runs through many of the concerns above: erosion through convenience. A practical concern arises when teams default to AI as a first step rather than using it to support thinking. Drafts are produced quickly, often lacking context or care, and treated as "good enough", narrowing opportunities for reflection.

A related concern is not sudden deskilling, but a gradual thinning of engagement. When AI removes friction too effectively, it can also remove the effort through which judgement is practised and confidence is built. Over time, people can gradually lose touch with their own voice.

These dynamics are unevenly felt. Differences in comfort with AI create new forms of team friction. Some people feel pressured to adopt tools that do not sit well with their professional identity. Others feel constrained by colleagues who resist AI use entirely. Without shared norms, these differences can harden into persistent tensions.

At the far edge of this concern sit anxieties about persuasion. When AI-generated insights are used to steer discussions or justify pre-set decisions, particularly without transparency, people feel their capacity for informed judgement has been undermined. In these moments, the issue is not efficiency but the legitimacy of collective decision-making.

Convenience becomes problematic when it replaces engagement rather than supporting it. In collective work, some forms of effort are not inefficiencies to be removed, but sources of meaning, learning, and trust.

Rather than asking how much AI is too much, teams may find it more helpful to ask where effort still matters. Writing, discussion, and reflection are practices through which people test ideas and surface disagreement. Protecting space for that work helps ensure that AI remains a support for judgement rather than a substitute for it.

Concluding thoughts

Taken together, these tensions point to a common underlying issue. AI does not introduce entirely new problems into collective work. Instead, it amplifies long-standing questions about authorship, responsibility, trust, and how judgement is exercised together. What feels unsettling is often not the presence of the technology itself, but the speed and subtlety with which it reshapes practices that were previously held through habit, craft, and relationship.

Seen this way, the current moment calls for greater deliberateness in practice, alongside whatever formal rules and safeguards the setting requires. In shared work, judgement has always depended on people having time and space to think together, to contest interpretations, and to stand behind what is produced. Used well, AI can sometimes widen that space, helping groups surface options, perspectives, or connections they might not otherwise reach. It can also make it easier to move quickly past the very moments where deliberation and shared sense-making still matter. Paying attention to discomfort can help teams notice when care, deliberation, or shared sense-making are being compressed too far.

Used with this awareness, AI does not need to become an authority in the room. But nor can we always assume that its role will be neatly contained or introduced through a single, visible route. People may increasingly bring AI-assisted preparation, translation, interpretation or drafting into shared work themselves, while AI may also shape the records and syntheses produced around that work. The task is therefore less about drawing a firm boundary around AI use than keeping enough of the process open to question: what contributions and interpretations are being taken to represent, where judgement is being exercised, and whether people can still recognise, contest and stand behind what emerges.

Holding that space is an ongoing practice, shaped by context, relationships, and the kind of work people are trying to do together. One place where this becomes especially visible is in the summaries, syntheses and draft framings that groups begin to work from.

4. AI and the authority of the first synthesis

Many teams, workshops and collaborative processes are starting to use AI to summarise notes, draft strategies, cluster themes, prepare options and review emerging ideas. These uses can be valuable, especially when they help people move from scattered material to something they can discuss. But an AI-assisted draft can also shape what a group notices, questions and treats as ready to work from. This essay looks at how teams can use AI-generated syntheses as working material, while keeping the framing open to collective judgement.

In collaborative work, the first organised account of an issue often matters more than we realise.

A facilitator groups workshop comments into themes. An evaluator synthesises interview findings. A programme team drafts a Theory of Change. A policy group turns scattered evidence into an options paper. These are working drafts, but they still carry influence.

A first synthesis gives people something to respond to. It also brings choices forward: the language used, the boundaries drawn, the way ideas are grouped, and the account of what may already be known. A set of themes suggests what belongs together. A diagram suggests what counts as a pathway, a boundary, a condition or an outcome.

This has always been part of facilitation, evaluation, planning and systems work. Generative AI does not create the issue, but it changes the conditions under which early syntheses are produced. AI can now summarise notes, cluster comments, draft a Theory of Change, compare options, produce a risk scan or prepare a discussion paper much more quickly than before. This can help people move from scattered material to something they can question, revise and discuss.

But it can also make early framing feel more complete than it is.

There is a further complication. Some of the material entering a collaborative process may already have been shaped with AI, for example through translation, preparation, drafting or help in developing a contribution. That does not necessarily make the contribution less valid, and in some settings it may help people participate more fully. But it can matter for how the material is interpreted. A considered position developed with assistance is not necessarily evidence of the same thing as an immediate response, unaided understanding or someone's own first framing of an experience. Before synthesising such material again, we may need to be clearer about what we are treating it as evidence of.

The issue, then, is not simply whether AI supports or replaces human judgement. It is how collective judgement is exercised when AI has helped prepare an early version of what people are asked to judge.

What AI changes

The practical change is speed and fluency.

AI can make early syntheses faster, smoother and easier to produce. That can be a real gain. It can create a starting point where people are struggling to get started, help a team compare different ways of grouping the same material, or reconnect people with ideas scattered across documents, notes and memories.

The risk is that fluent first drafts can feel more settled than they are.

The categories can arrive already named. A synthesis may be broadly accurate while still narrowing later discussion, smoothing over disagreement, or missing what people know but have not written down.

When a working draft becomes the frame

This becomes sharper when an AI-assisted synthesis becomes a shared object for a group.
The concern is not only that one person may be influenced by an AI output. It is that a group may begin to organise its shared attention around an artefact that no one in the group has fully authored, and that no one has yet tested carefully enough.

A working draft can then become more than a draft. It shapes the questions people ask, the disagreements they notice, the pathways they consider, and the decisions that seem ready to be made. This concern has a parallel in research on anchoring in AI-assisted decision-making. Carter and Liu (2025), in experiments with managers making performance judgements, found that AI recommendations could influence subsequent ratings. Their setting is different from collaborative synthesis, but the underlying caution is relevant here: an early AI-generated contribution can become an anchor for the judgement that follows.

A collaborative strategy process provided one practical example of how this can work in both useful and potentially problematic ways. A small team was developing a strategy for a complex area of practice. The group knew the work well, but the discussion stayed close to how the work was already organised. It was harder for them to step back and articulate the wider strategy that might be needed.

To broaden the discussion, I asked an AI tool to prepare a first draft strategy, using material that was suitable for that purpose, rather than confidential notes or unreviewed stakeholder input. The draft was not context-rich, and it was not ready to use. But it pulled together a wider set of functions, dependencies and practical requirements than the group had been able to name from within the existing conversation.

The draft was then brought back to the team with its AI provenance made explicit. That disclosure mattered. Without a visible author, it was less clear whose lens the draft carried. Once its provenance was clear, they seemed better able to treat it as something to be tested rather than as an authoritative account.

The draft was more useful than expected, not because it was right, but because it opened up the discussion. The team then changed it substantially: correcting the language, adding context, removing what did not fit, and reshaping it around their own understanding of the work.

The example shows both the value and the risk. The AI draft helped unlock the conversation. But it only became useful because people treated it as provisional and reworked it carefully. In another setting,

with less time or support for review, the same kind of fluent first draft could easily have been accepted too quickly.

This does not mean AI should not be used to help prepare material for shared work. It does mean that the first AI-assisted synthesis needs to be treated as something to examine, not simply something to improve.

What needs to stay open

The issue is not only what AI can or cannot do. It is what people allow AI-shaped material to frame before they have had a chance to examine it together.

Some things AI cannot hold in practice. It can process text about relationships, place, history, culture, trust and lived experience, but that is not the same as carrying those things in the work itself. It does not hold the relationships, or share the accountability. It does not know what it means for a group to work together over time in a particular setting.

Other things AI may be able to influence, but should not be allowed to define. It should not determine who has standing in a conversation, whose knowledge counts, what trade-offs are acceptable, or when enough agreement has been reached. These are questions of legitimacy, responsibility and judgement. There is also a middle space that needs care. AI can work with codified local knowledge: reports, plans, transcripts, submissions and datasets. It can help reconnect people with material that is too easily lost across long processes.

But codified knowledge is not the same as knowledge held in practice. A report may contain a community's words, but not the relationships, histories or tensions that shaped those words. A transcript may preserve what was said, but not always what was difficult to say. In a long-running place-based programme, for example, a report may record what people said, but not the trust, caution or history that shaped how they said it.

So the task is not simply to ask whether the synthesis is correct. We also need to ask what kind of account has been produced: what has been grouped together, what has been left unnamed, and what someone who was present might recognise as missing.

In shared work, a first synthesis should help people enter the conversation without setting its terms too early.

Collective judgement as a practice

Collective judgement is not just the final decision a group reaches. It is the shared work of noticing what matters, questioning how things have been framed, and deciding what needs attention.

AI can support this work when it helps people see more, prepare better, or test their thinking. But it can weaken the work when it makes a particular framing feel more settled than it should be, encouraging premature closure around a version that still needs to be questioned. Keeping a human in the loop is not enough if the process itself no longer supports questioning, challenge and shared interpretation. Human judgement may still be present, while the group's capacity to exercise judgement together gradually weakens.

This is why the practice around AI matters as much as the output.

If AI has helped prepare a summary, diagram, briefing or draft framing, its role should be visible enough to discuss. People need to know what kind of artefact they are working with. That helps them decide how much confidence to place in the material, and how to question it.

There is also a facilitation task here. A group may need clear permission to challenge the synthesis rather than simply improve it. It may be useful to ask people what feels missing, what feels too tidy, what wording does not sit right, or what another group might say differently. The point is to reopen the frame before it becomes too comfortable.

This is not about slowing every process down. But where AI has helped create an early synthesis, part of the work is to make sure the group can still judge the synthesis itself.

Questions to keep the frame open

The following questions may help teams, facilitators, evaluators and programme leads work with AI-assisted syntheses more carefully.

- How has AI been used in preparing this material, and what material did it work from?
- What are the contributions being treated as evidence of, and does how they were produced matter for that interpretation?
- What has the synthesis grouped, renamed, smoothed or made more prominent?
- What assumptions does this version make easier to accept?
- What forms of knowledge, context or disagreement may sit outside the version now in front of us?
- Who has reviewed the framing, and what still needs to be questioned before this becomes the version people work from?

The point of these questions is not to create a new compliance process. It is to keep the work of judgement visible. A first synthesis is useful when it helps people think together. It becomes risky when it quietly closes down too much before people have had a chance to examine it.

Closing thoughts

The ease of producing an early synthesis changes the practice around it.

When a draft appears quickly, clearly and confidently, people may move too soon to improving it rather than questioning how it has framed the issue. They may treat it as a neutral starting point, especially when it arrives without a visible author whose angle they can read.

The challenge is not to keep AI away from collaborative work, or to pretend that human syntheses are neutral. What matters is making the work of synthesis visible enough that people can question it together.

AI may help prepare the ground, but the work of judgement still needs to be held by people who can ask what this draft has made easier to see, what it has pushed to the side, and what the group still needs to work through. The same distinction matters when the resulting record is taken as evidence of what participation itself produced.

5. AI can produce the record of participation without the participation

Organisations evidence participation through its visible products: workshops held, submissions received, themes identified, reports produced. Much of the documentary record around those activities can now be generated in minutes. That makes it possible to produce a convincing account of a participatory process while a good deal of the process itself has quietly gone missing, and nothing in the reporting will show the difference. The more important question is what participation was doing that the record was never able to capture.

Much of that record can now be generated. AI can transcribe a discussion, sort comments, identify themes and produce a summary that reads as though a group worked its way towards something. The artefacts look much as they always did, because they were always the visible part: the part that could be counted, inspected and filed.

It's worth being precise here. The artefacts were evidence of something. The question is what, and what happens when they can be produced without it.

The gains are real

Engagement has often placed unreasonable burdens on people, and those burdens have fallen most heavily on those with the least time and resources. Attending three evening meetings, reading a hundred-page discussion document, and writing a submission in a register that officials will take seriously are all forms of unpaid work.

AI can genuinely reduce some of this. It can make asynchronous and recorded contributions easier to translate, summarise and work with, while plain-language tools can make material accessible to more people.

The effort was never evenly shared, which complicates any argument for preserving it.

Nor does every process need to involve shared framing or collective decision-making. Much statutory consultation is legitimately about gathering responses within a decision frame that has already been set, and I have argued elsewhere that good engagement does not mean involving everyone in everything. Where that is the honest purpose, doing it faster and more accessibly is often better, though it still leaves questions about how responses are categorised and represented.

There is another complication here. AI may increasingly be used by participants themselves, to understand material, translate it, prepare for a discussion, organise their thoughts or help put a contribution into a form they feel will be heard. That does not necessarily weaken participation. Some of these uses may reduce barriers that participatory processes have always created for people who are less confident with formal language, working in another language, short of time, or unfamiliar with professional ways of presenting an argument. But it makes clarity about the purpose of the activity

more important. A contribution developed with AI may be perfectly good evidence of someone's considered view, while being much weaker evidence of their unaided understanding, their first response, or the way they themselves initially framed an experience. The issue is less whether AI was involved than what we are treating the contribution as evidence of.

The distinction becomes more important when a process was meant to do more.

When participation is intended to do more than gather responses, the difference between supporting the process with AI and producing its visible record without the process becomes important.

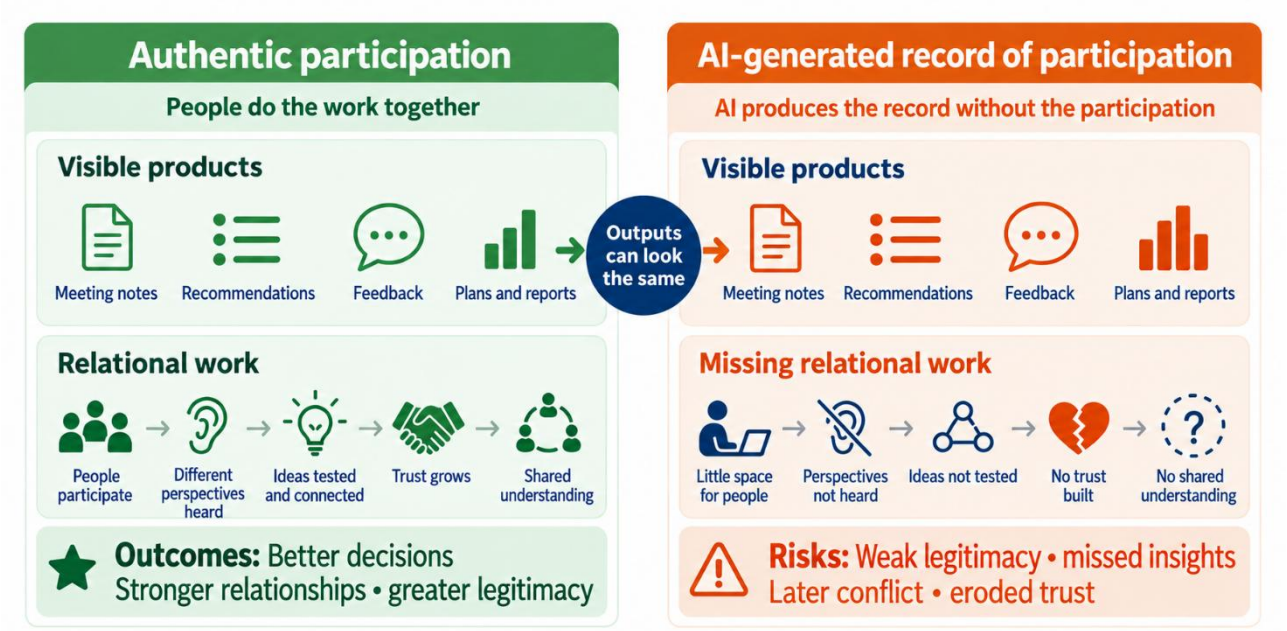

Figure 3: Using AI with and without participation. The visible products can look similar, while the relational work beneath them may be very different.

The figure separates the visible products of participation from the less visible work through which people influence how an issue is understood, hear different perspectives, test interpretations, build trust and develop shared understanding. AI can help capture, organise and analyse what people create together. But when the same kinds of outputs are produced without that participatory work, the record may look complete while much of what gave the process value is missing.

What the process was doing

Participation can allow people to influence how an issue is understood, hear and respond to one another, test interpretations, work through disagreement and develop some ownership of what happens next. It can build relationships and shared judgement that stay useful long after the report is filed. Those outcomes are why the effort was worth spending.

Much of this is difficult to see in the record alone. A summary may not show whether participants influenced the framing or merely commented on someone else's. Themes may not show whether a disagreement was worked through or smoothed over. A list of workshops does not tell us whether people left understanding how a decision would be made and who would make it.

We have relied on the artefacts as a proxy because the things that mattered were hard to evidence and the products were easy to count. The proxy was never reliable. Arnstein (1969) made a related point years ago: consultation can create the appearance of participation without giving people meaningful influence, and institutions have long been able to report participatory activity without changing who actually shapes decisions. Much of that was never deliberate. It was habit, procedure, and the momentum of a process that had to be seen to happen.

More recent work on participatory approaches to AI has raised a related concern. Sloane and colleagues (2022), for example, warn against "participation-washing", where the presence of participatory activities does not necessarily mean that those involved have meaningful influence.

What has shifted is how much easier the gap is to open. Producing a convincing record used to require sustained contact with the material even where the engagement behind it was thin. That contact is now optional. An organisation acting in good faith can meet every reporting requirement, produce material of a better standard than it has managed before, and see nothing in its own reporting that would suggest anything was missing.

> **A good record of a collaborative process is not the same as a collaborative process, and it has become considerably easier to have one without the other**

Which effort was carrying the work

As a facilitator I have an obvious interest in arguing that process matters, and the argument should not rest on that. Activities should not survive because facilitators have traditionally done them, and nothing here requires collaborative work to stay labour-intensive.

The question is which effort was carrying part of the work, and it is harder to answer than it first appears. Some of it was clerical. Reformatting a document for a third audience, chasing attendance lists, tidying a table: reducing that is a plain gain.

But the obvious candidates are not always the safe ones. Transcription is what most people cut first, and qualitative researchers have argued for decades that transcribing is where you come to know your material. Listening twice, deciding how to render a hesitation, noticing on the second pass what you missed on the first: that is analysis, not typing. When AI does it, the text arrives without the immersion that used to come with producing it. The time is genuinely saved, and something now has to replace what the time was also doing.

Other effort was more obviously doing the work. Comparing two readings of the same material and finding they do not agree. Struggling to name an outcome and discovering in the struggle that the group means different things by it. Returning to a disagreement that was parked three meetings ago because it has surfaced again in a different form. These look inefficient, and they are often where the understanding actually forms.

Holding several readings at once

There is a way of describing this that I have found useful.

In a long collaborative process I am rarely reading one thing. I am reading the people, and whether they are anxious, tired or engaged. I am reading the state of the task, and whether we are generating enough or converging too early. I am reading the maturity of the process itself, and whether the design is still doing what we set it up to do. I am reading the trajectory towards whatever we agreed we were working on, and whether we have drifted. And I am reading the wider context: the relationships, history and commitments outside the room that will determine whether any of this holds.

Those readings frequently disagree with one another. The energy in the room says stop while the funding timeline says push on. The content is converging nicely, and converging on the wrong problem. The group is ready to decide something the wider context is not ready to receive.

Holding those readings open, rather than resolving them too early, is a big part of the work. The judgement lies in the conflicts, not in any single reading.

Producing an account usually means collapsing them into one version, arrived at by deciding which readings governed. It does not have to. An account can carry competing interpretations, name what remains unresolved, and say which questions a group could not answer. Those accounts are harder to write and harder to commission, and they are rarer than they should be. The skill is knowing when to collapse them, and when to leave them open until the group is ready to make that judgement and own it, which is the argument developed in Essay 4, *AI and the authority of the first synthesis*.

An AI synthesis can only work with what has been made legible to it. Some of the readings can be supplied. You can describe the history, hand over earlier reports, say where the process has drifted, and ask explicitly for contradictions to be kept rather than resolved. Most of us do less of that than we should.

But much of what is being read never enters the record in the first place. It is embodied, provisional, held tacitly between people who have worked together for two years, or noticed and not yet put into words. It was never in the transcript to be handed over. An account can therefore read as coherent partly because tensions that were live in the process were never present to the system as tensions. The output is coherent because, as far as the system could tell, nothing was in conflict.

Where this leaves AI policy

Many organisational responses to AI sit with information technology, legal, privacy or risk teams, and they ask whether a use is permitted, secure and appropriately disclosed. All of that matters. On its own it is not well placed to notice what this essay describes.

A consultation report can be produced in a compliant environment, clearly labelled as AI-assisted, meet every privacy requirement, and still leave participants with no way to challenge how their contributions were categorised. Nothing prevents a governance framework from asking about contestability, participant review, or how an interpretation was arrived at, and it is worth looking at whether yours does.

The distinction to look for is between oversight of an output and attention to the process that produced it. That attention may increasingly need to extend to how contributions themselves are produced, but without turning transparency into surveillance. The relevant question is not whether every use of AI can be identified, but whether we know enough about the process to interpret what people contribute appropriately.

The practical response therefore sits largely with the people designing the processes. It might mean delaying the first synthesis until participants have developed their own reading of the material. It might mean asking AI for three possible interpretations rather than one, and putting all three in front of the group. It might mean deliberately preserving disagreement instead of resolving it, keeping source material accessible so people can check what happened to their contribution, or inviting participants to revise the categories rather than only comment on the summary.

It also means being honest about where the time goes. If AI saves a facilitator two days, those two days can be absorbed as efficiency or spent on interpretation and dialogue. The second does not happen by itself, and ordinary reporting will not register which one occurred.

What we now have to say out loud

The uncomfortable implication is that we can no longer let the record stand in for the process. It never fully could, and the gap has widened.

That means saying what participation is for in a given piece of work, before it starts, and being willing to report on whether it did that rather than on how many workshops were held. It is harder than counting. It is also what now separates a process that genuinely achieved its participatory purpose from one that only sounds as though it did.

AI will keep getting better at producing the outputs of collaboration. We will have to get clearer about what the collaboration itself was for.

PART 3
Learning, capability and organisational practice

Using AI well depends not only on individual judgement, but on how people and organisations learn from experience. This section looks at how we can judge whether AI is actually improving the work, how teams and organisations can begin developing shared approaches, and what AI-supported practice means for education, professional learning and capability over time. It also returns to place-based practice, where these questions are shaped by context, relationships and long-term responsibility.

6. How do we know whether AI is improving the work?

Generative AI is increasingly being used in products, services and professional work, but evaluating whether it is helping requires more than checking its outputs. This essay draws on established approaches to planning, indicators, evaluative rubrics and monitoring, evaluation and learning. A composite consultation example shows how an AI-supported output can be followed into the work and decisions that come next.

In this essay, evaluation means deciding what good use of AI would look like, gathering evidence about what actually happens, and using that evidence to improve the work. In AI development, structured ways of making these checks are often called "evals".

Some AI evals focus on whether a system produced an accurate answer, followed its instructions or completed a defined task successfully. Test sets, reference material and rubric-based grading can provide useful evidence about factual support, relevance, completeness and other qualities of an output. AI eval tools can also include model-based graders that assign scores or labels against specified criteria. These checks matter. An inaccurate answer, unsupported analysis or misleading summary is unlikely to help anyone do better work.

But output performance is only part of the evaluative picture. Weidinger and colleagues (2025) argue for a more developed evaluation science for generative AI, including measures that relate to real-world performance, sustained human–AI interaction and broader contextual factors. That wider view is particularly important when the question is not simply whether an AI system performed well, but whether its use actually improved the work.

Whether AI is used by an individual practitioner or built into a product, service or organisational process, a satisfactory result at the point of generation may not show what changed around it. An AI-assisted report may be factually sound while giving readers too much confidence in an uncertain conclusion. A product feature may produce accurate responses while leaving users unclear about its limitations or how much reliance to place on it. An AI-supported service may reduce response times while making unusual or higher-risk cases harder to identify and escalate.

A satisfactory output does not show, by itself, that the work improved.

These are not entirely new evaluative problems. Approaches to planning, indicators, rubrics and monitoring, evaluation and learning have developed over decades in fields dealing with contested purposes, partial evidence and different legitimate views of what matters. Generative AI introduces new technical possibilities and risks, but it does not require us to start again. These traditions can help

clarify what an AI use is meant to contribute, identify relevant evidence, support judgement and improve the practice.

The diagram below brings these elements together: what the AI use is meant to contribute, evidence about what happens in practice, informed judgement about what that evidence means, and decisions about what should change.

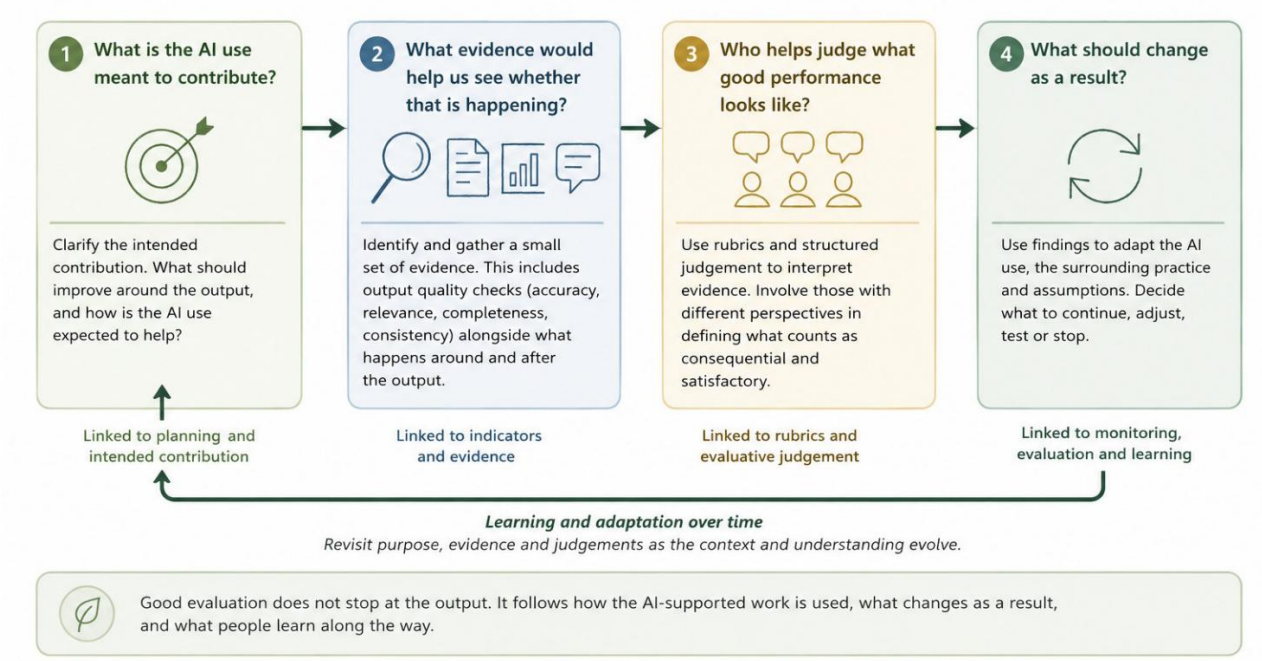

Figure 4: Evaluating AI-supported work as a connected process. Output checks form part of the evidence, but evaluation also considers the surrounding practice, what changes and what is learned.

Purpose, evidence, judgement and adaptation inform one another as the AI use and the surrounding work evolve. Output checks sit within this wider process, alongside attention to how the output is used, who helps interpret the evidence and what is changed as a result.

Start with what the use is meant to contribute

Evaluation begins before measures are selected or outputs tested. It begins with knowing what we are trying to achieve and how the proposed use of AI is expected to help.

Clarifying the intended contribution means asking what should improve around the output. What the system produces is only part of that. An AI-enabled product feature may be intended to help users complete a task while understanding the limits of the response and knowing when other support is needed. An AI-supported service may aim to reduce response times without making unusual or higher-risk cases harder to recognise and escalate. In professional work, the intended contribution may be to retain conflicting evidence or reduce routine effort while continuing to support thinking and judgement.

These descriptions provide a reference point for evaluation. Without that clarity, it is easy to measure what is readily available: user numbers, task-completion rates, response times, output volumes or reported time savings. Such evidence may tell us about uptake, activity and efficiency without

establishing that the product, service or work has become better. It may also leave assumptions untested, such as whether users recognise uncertainty or human review recovers material softened during automated synthesis.

This is where planning and monitoring, evaluation and learning need to remain connected. Planning clarifies where we are trying to go and why we expect AI to help. Monitoring and evaluation provide evidence about what is happening, while learning enables people to revisit their assumptions and adjust the practice.

Move from evidence to judgement

Once the intended contribution is clear, we can consider what evidence would help us judge whether it is occurring.

Indicators can help organise that evidence, but they remain partial and need to be interpreted in relation to purpose and context. Faster completion does not necessarily mean a better service. Increased use may show that a feature is convenient without demonstrating that it is valuable. A high accuracy score may be reassuring without showing whether uncertainty was handled appropriately or people placed more confidence in an answer than the evidence justified.

A small amount of well-chosen evidence may be enough to help people understand what is happening and make decisions about the use. Indicators need to be connected to purpose, interpreted alongside other evidence and used to support learning and adaptation. They do not provide conclusions on their own.

Some aspects of performance can be measured directly. Others require judgement. A technically accurate response may still handle uncertainty poorly, while a faster service may become less responsive to people whose circumstances do not fit common patterns.

Rubrics are already used in AI evaluation to judge outputs against stated criteria. The further question is what the rubric is designed to evaluate. Where the intended contribution concerns a wider product, service or practice, criteria may need to address whether users understand the limits of a response or unusual cases reach appropriate human support.

Developing such criteria requires more than extending an output-scoring rubric. Teams need to examine real cases and clarify what good performance would mean in the wider setting. Criteria and descriptions of quality can be developed with the people involved and refined as understanding grows.

Together, indicators and rubrics help extend evaluation beyond the immediate output. Indicators point towards evidence of what is happening in the product, service or practice. Rubrics help people judge what that evidence means in relation to the contribution they were seeking.

Follow an issue beyond the output

One way to evaluate an AI-supported practice beyond the immediate output is to follow an important issue through the workflow. This can show whether it remains visible as the output is reviewed, used to frame later discussion and carried into decisions. The following composite example, drawn from

recurring patterns in facilitation, synthesis and review and adapted to an AI-supported process, illustrates this approach. It does not describe a single project.

Suppose an AI-generated consultation summary is clear, concise and broadly accurate. It captures the main themes and areas of agreement, with few factual errors. One disagreement, however, is reduced to a short qualification within a larger theme. The summary is then used to prepare the next meeting, where the disagreement does not appear on the agenda, is not discussed further and is absent from the eventual decision record.

An output-level assessment might reasonably conclude that the summary performed well. To assess whether its use supported the wider work, we would also need to follow what happened to the disagreement. We could examine how it appeared in the original material, how it was represented in the synthesis and whether it remained available for later consideration. Much of the evidence may already exist in the notes, generated summary, human revisions, meeting agenda and decision record.

Tracing the issue would not prove that AI caused it to disappear. Human-written summaries have always selected, condensed and sometimes flattened what people said. The original notes may not have captured the disagreement clearly. The AI summary may have softened it, while a human reviewer chose not to restore it. It may have remained visible in the final summary but been omitted from the next agenda. Each finding points towards a different part of the work requiring attention.

The trace might also show that the disagreement remained clearly represented, reached the next discussion and was visibly considered in the eventual decision. That is equally useful evidence. The purpose is not to assume that something has gone wrong, but to understand where important issues are being retained, altered or lost.

The same approach can be adapted to other intended contributions. A service organisation, for example, might follow an unusual case through automated triage, staff review and final resolution to see whether it was recognised and escalated appropriately.

Looking across several consultation cases might eventually support a provisional description such as:

Consequential differences remain visible enough to be understood and considered at the next relevant stage.

This could become a reference point within an evaluative rubric, but only after being tested and discussed in the setting where it would be used.

Who decides what good looks like?

Once evaluation moves beyond the immediate output, the question of who helps define good performance becomes central.

A developer might judge the consultation summary by its factual accuracy, adherence to instructions and consistency of format. A facilitator may notice whether an unresolved issue has been absorbed into a majority theme. Participants may care not only that their views appear in the document, but whether consequential concerns remain available for later consideration. Decision-makers may need to know where conclusions remain contested and how those differences have been handled. Each sees a different part of what good performance involves.

What counts as consequential cannot be determined by frequency alone. A view may matter because it identifies a serious risk, represents people likely to bear the consequences of a decision, challenges a central assumption or draws on knowledge that others do not hold.

Asking people to correct a completed summary may reveal particular omissions. Involving them in deciding what the evaluation should look for may reveal a different set of criteria altogether. This need not become an elaborate exercise. A small trial might involve reviewing a few routine and difficult cases with people who bring different perspectives, then agreeing on one or two provisional criteria for the next round.

Use evaluation to improve the practice

The point of evaluating generative AI use is not simply to determine whether a tool, feature or workflow passes or fails. Findings may lead to changes in an interface, explanations of uncertainty, routes to human support, escalation procedures, review responsibilities or the circumstances in which AI is used. They may also lead people to narrow the intended use, reconsider the assumptions behind it or decide that AI is not appropriate for a particular task.

This is what it looks like when evaluation feeds back into practice rather than ending with a verdict. Evidence is interpreted and used to change the tool, the surrounding work and sometimes the decision to use generative AI at all.

The same principle applies across an AI-enabled product, service or professional practice. We need to connect purpose, evidence, judgement and learning closely enough to see whether using AI actually improved the work.

7. Working with AI: where to begin?

Generative AI raises questions about usefulness, ethics, judgement and responsibility. For organisations and teams, there is also a more immediate question: where do we begin if we want to engage with AI more deliberately? A useful starting point is to look at what is already happening rather than wait for a complete policy or strategy.

This practical emphasis is consistent with wider work on operationalising AI ethics. Morley and colleagues (2023), writing mainly about AI development, found that putting ethical principles into practice depends on practical support, clearer responsibilities and organisational conditions that allow people to raise and work through concerns.

Seeing the overall shape

When organisations begin working with AI, a few areas usually come into view:

- who is paying attention and providing direction
- how teams are making sense of where AI could help and where it could cause problems
- what shared expectations about use and limits are needed
- how the organisation keeps learning as things change

These areas can be considered across an organisation, within a programme or by a particular team. The scale may differ, but the conversations are similar. They do not need to be addressed all at once, or in a fixed order. Together, they help show the overall shape of the work. If you are just getting started, it is often enough to pick one or two of these areas and begin with a small conversation. You do not need a full plan in place.

1. Getting the right people engaged

This is often where things start. Leadership takes an active interest in how AI is already being used, and brings together a small group to look at it from different angles. In many organisations, AI tools are already in use informally. Starting with what is happening, rather than what should happen, tends to ground the conversation. In many cases, this can be as simple as a short discussion with a few people who are already using these tools.

A question to start with: Where are we already using AI, even informally, and who needs to be involved in thinking about it?

Things worth exploring:

- What are people using AI for, drafting, analysis, summaries, something else?
- What has felt useful, and where has it felt uncertain?
- Who understands the work context well enough to see where AI matters? Who understands the risks or obligations? Who is already experimenting?

Getting these people into the same conversation early makes a difference. It does not need to be a formal committee. A small working group with a clear brief is usually enough.

2. Making sense of AI with teams

Once there is some leadership attention, the next step is often to create space for teams to think about what AI means for their own work. This matters because AI tends to show up in specific tasks, decisions, and relationships. The people doing the work are usually best placed to see where it could help, where it could cause confusion, and where they would want to keep human judgement in place.

A *question to start with*: How might AI change how we do our work, for better or worse?

Things worth exploring:

- Where are the time pressures or bottlenecks where AI might be genuinely useful?
- Where do we rely on interpretation, relationship, or contextual judgement?
- What might change, in the work itself, or in how it is received, if AI were introduced?

These conversations often surface concerns that would not come through in a top-down process. They also tend to build a more realistic picture of where AI actually fits.

3. Clarifying boundaries and developing shared guidance

As the picture becomes clearer, most organisations find they need some shared expectations about how AI is used. This does not have to be a formal policy from day one. But it does mean working through a few key questions together.

The harder conversations tend to sit here. Questions about what is acceptable to share with an AI tool, what needs to be checked before use, where the organisation would draw a line, and what kinds of risks need to be managed are not always straightforward. Different parts of the organisation may see these differently, and that is worth surfacing rather than trying to resolve everything at once.

A *question to start with*: What would good use and poor use of AI look like in our context?

Things worth exploring:

- What would we be comfortable with if it were visible externally?
- What should be reviewed or checked before use?
- Who is responsible for AI-assisted outputs?
- When should AI use be disclosed?
- What kinds of risks are we most concerned about in our context?
- What do people need to know to use AI responsibly here?

The aim is to capture these decisions in a form people can actually refer to: something light and usable rather than a document that sits unread. It will need revisiting as experience grows.

4. Regular review and learning over time

AI tools and their uses will keep changing, and so will the organisation's understanding of where they help and where they do not. Building in some form of regular review, even lightweight, helps the organisation stay responsive rather than locked into early assumptions.

This is also where the quality of organisational learning shows. It is relatively easy to adopt AI tools. It is harder to notice when something is not working well, when a boundary needs adjusting, or when early enthusiasm has outrun careful use.

A *question to start with:* What are we noticing from using AI, and what might we change?

Things worth exploring:

- What has worked better than expected?
- What has not sat right, and why?
- What would we want to adjust, in how we use AI, or in the guidance around it?
- Are there new uses emerging that we had not anticipated?

Over time, this kind of review helps teams see more clearly where AI is useful, where problems are emerging and what needs to change.

Closing reflection

Organisations do not need a complete system in place before they begin. Shared understanding can develop through use, discussion and review.

In many settings, the quality of judgement around AI, how it is used, interpreted, and questioned, matters as much as the technology itself. These starting points are simply a way to begin while keeping learning and context in view. Over time, that also becomes a question of how people develop and maintain the capability to use AI well.

8. Rethinking education, training and professional learning for AI-supported practice

As AI increasingly becomes part of professional practice, we may need to rethink not only how people use it, but how people learn to become, and remain, capable practitioners.

Essay 1, *What would better work with AI look like?* began from the quality of the work itself. The related question here is what people need to learn, practise and retain as AI becomes part of that work.

Much of the current discussion about AI and education starts with cheating. That is understandable. Schools and universities need to know whether submitted work represents a student's own effort, and qualifications need to mean something. But there is a risk that we hold the existing educational model constant. We continue to set much the same essay, assignment or task, then focus on how to prevent students using AI to complete it.

That may address an immediate problem without getting to the larger one. If AI is going to be part of the work people eventually do, should we also reconsider the learning goals, activities and assessments themselves?

This is a different question from how AI might improve existing education. AI may provide better tutoring, faster feedback, help with drafting, or easier access to information. All of those may be useful. But they largely ask how AI can help us do familiar forms of education and training better.

A more fundamental question is what people need to learn differently when AI becomes part of the practice for which they are being prepared.

The distinction matters well beyond schools and universities. AI is becoming part of research, policy, medicine, engineering, evaluation, law, teaching and many other kinds of professional work. It is increasingly used to search, draft, summarise, compare, analyse and suggest possible interpretations. Professional development is changing too, because experienced practitioners are now learning while working alongside systems that can take on parts of activities through which they once developed and maintained their expertise.

There is an important difference between being able to use AI and being capable of using it well.

A person can learn how to prompt a system, ask for alternatives, request a summary or generate a first draft quite quickly. But using AI well depends on much more than operating the tool or checking its outputs. It depends on understanding the work itself: what matters, what good work looks like, what

can reasonably be inferred, where context matters, and when further inquiry, judgement or discussion is needed. That wider understanding is what allows someone to decide where AI can help, where it may mislead, and where it should remain only one part of the process.

Those capabilities do not appear simply because AI is available. They have to be developed somewhere. This connects with work by Bearman and colleagues (2024) on evaluative judgement, including how people learn to judge both generative AI outputs and the processes used to produce them.

That raises questions about what people still need to understand, practise and experience for themselves. Some activities that look productive on the surface are also doing developmental work at the same time.

When doing the work is also part of the learning

We write a report because a report is needed. We analyse a set of data because we need the result. We discuss a difficult case because a decision has to be made. But these activities often do two things at once. They produce an output, and they develop the people doing the work.

Writing can help us discover what we think. Comparing several cases can teach us which differences matter. Struggling with an analysis can reveal where our understanding is weak. Discussing an interpretation with colleagues can expose assumptions that would otherwise remain invisible. Supervising someone through a difficult piece of work can develop the judgement of both people involved.

What makes these activities developmental, though, is not always the task itself. Feedback, having to explain a judgement, encountering disagreement, or later discovering whether we were right can all matter. So preserving the activity may not be enough. We need to understand the conditions that make it developmental.

Suppose a multi-site evaluation has generated findings from a number of local projects. At some point those findings need to be brought together into a smaller number of wider conclusions. AI can help organise material, compare reports, identify recurring themes and suggest possible patterns.

That could save considerable time. It might also improve the process by making it easier to work across a large body of material.

> **When AI takes over part of the work, we also need to ask what learning that work was carrying.**

But much of the judgement lies in making sense of that synthesis. Are two findings really describing the same thing? Is a difference between sites incidental, or does it change the meaning of the finding? Can something reasonably be generalised, or is it too dependent on context? Does an apparent pattern reflect the evidence, or simply the way the material has been framed?

As discussed in Essay 4, *AI and the authority of the first synthesis*, an early AI-generated interpretation can also quickly shape what is discussed next. Those questions therefore involve more than checking whether the synthesis is accurate. Some of the judgement is developed and exercised through the process of working them out together.

A group may begin with different interpretations and gradually come to understand why they differ. Someone may notice that a conclusion that seems obvious from one perspective looks quite different from another. A discussion may reveal an assumption embedded in the original questions. The learning is not separate from the work. It happens through doing the work, and sometimes through having to make sense of it with others.

If AI provides an early, plausible synthesis, that may be useful. But it may also shape the discussion that follows. Instead of asking openly what the material might mean, people may find themselves reacting to categories and patterns that have already been proposed.

The issue is therefore not whether we should use AI for synthesis. It is whether we understand what else was happening in the activity that AI is now helping us perform.

That question matters for curriculum design, training and workplace learning. If an activity once produced both an output and a learning experience, we need to know whether handing more of it to AI changes the learning as well as the workload. If it does, we then need to ask whether that learning still matters and, if it does, where it will happen instead.

Developing judgement, and keeping it

This is not only a question about novices.

Experienced practitioners may be particularly good at working with AI because they have years of practice against which to judge what it produces. They have seen enough cases, mistakes, exceptions and awkward situations to recognise when something does not quite fit.

There is a transition problem here. Many people who are good at judging AI-supported work developed that judgement through years of doing work that AI may now partly take over. Where will the next generation get the experience on which that judgement depends?

> **Judgement is not something we accumulate and then store away. It has to be exercised, tested and renewed as the practice itself changes.**

If AI increasingly takes over parts of the activities through which practitioners encounter difficult cases, compare alternatives, work through uncertainty or explain their reasoning to others, what happens to that judgement over time?

Developing judgement and maintaining it are not quite the same problem. A novice needs opportunities to build it. An experienced practitioner needs opportunities to keep using it.

This does not mean people should continue doing everything manually. Some activities may no longer deserve the time they once consumed. AI may remove routine work that was never especially developmental in the first place. It may also create new learning opportunities, for example by offering rapid feedback, generating alternative explanations or making it easier to explore several approaches to a problem.

Working with AI can itself be developmental if people have to form their own view, compare it with what AI produces, and explain where and why they differ. The more useful question is which experiences matter for capability, and why. This is also a question of human agency: whether AI strengthens people's capacity to understand, judge and act, or gradually reduces it. Essay 11, *Designing AI for the wider world: six principles for better development*, considers what these concerns might mean for better AI design and development.

Where does judgement move?

That also changes how we think about keeping a human in the loop. The phrase can make it sound as though judgement remains in the same place and a person simply checks the machine's work before it moves on.

In practice, judgement may move.

It may move to the person who frames the initial question. It may be shaped partly by assumptions embedded in the system, and by the categories or patterns the system makes readily available. It may move downstream to whoever decides whether an output is acceptable. Or a judgement that once emerged through discussion among several people may shift into an individual interaction between one person and an AI system.

In the synthesis example, this could mean judgement shifting from a group working through interpretations together to one person framing an AI request and deciding whether its resulting synthesis is good enough.

That change matters for learning as much as it does for accountability, and it may also matter for collective sense-making. A process that still contains human judgement is not necessarily the same process if that judgement is now exercised individually rather than worked through with others.

> **Keeping a human in the loop does not necessarily keep judgement in the same place.**

For schools and universities, this suggests looking beyond whether students used AI. We also need to ask what a task was intended to develop, whether that learning still matters, and whether the same task remains the best way to develop it.

For professional preparation, the question is somewhat different. Where should people learn to work with AI, and where do they still need direct experience of the underlying practice because important capabilities depend on that experience? For continuing professional development, the focus shifts again. The question may be less about courses than about the work itself. Where will experienced

practitioners continue to encounter uncertainty, compare interpretations, discuss difficult cases, test their reasoning and learn from each other as AI takes on more of the routine work?

There is unlikely to be one model for this across different fields. What a surgeon needs to learn through practice is different from what a policy analyst, teacher, engineer or evaluator needs to learn. Nor will every activity that once had educational value need to be preserved.

But a small set of questions may help us think about what needs to change.

- What capabilities do we want people to develop and maintain?
- What do they still need to practise or experience for themselves?
- Which parts of the work can AI usefully take over?
- If an important learning experience changes or disappears, where will that learning happen instead?
- And where does judgement move when AI becomes part of the process?

If AI is changing professional practice, then improving existing education with AI is only part of the task. We may also need to work backwards from changing practice and ask what forms of education, training and professional learning will help people become, and remain, capable practitioners within it.

9. AI in place-based practice: what is shifting

Conversations about AI futures often sit at a distance from practice, focusing on policy, risk, or speculation. This essay begins from the ground: from the everyday work of evaluation, facilitation, and collaborative decision-making in place-based settings, where the issues at stake are often environmental, long-term, and contested. Drawing on futures thinking, it reflects on what seems to be changing as AI enters this work, what remains uncertain, and what others may be noticing in their own practice.

In place-based settings, relationships and responsibilities persist beyond any single programme. The work is often about real places and long-term consequences, where different perspectives and interests need to be worked through together.

Related work on community-based AI points in a similar direction. Hsu and colleagues (2022), drawing on case studies of AI developed with local communities, emphasise the importance of co-creation, local context and adapting systems to changing social conditions.

AI is beginning to shape how people prepare, form ideas and make sense of what to do next. Drawing on futures thinking, not as prediction but as a way of noticing patterns of change, I have been reflecting on what seems to be moving, what feels more established and what is only beginning to appear.

Noticing what is shifting

One way I've found helpful is to think in terms of movement across practice. Not as a formal method, but as a way of noticing what ways of working are receding, which pressures or drivers are becoming more prominent, and what new possibilities, or risks, are just beginning to come into view.

Some assumptions that seemed stable not long ago are starting to loosen. AI is increasingly becoming part of ordinary professional practice rather than something sitting at its edges. In many settings, it is already part of how people prepare, draft, and reflect, whether this is named explicitly or not.

Questions about authorship, accountability, and transparency are becoming more visible, particularly in shared documents and collaborative processes. Teams are beginning to ask how AI-generated material should be treated, and where responsibility sits when outputs are collectively produced but partly generated.

Alongside this, expectations around speed and polish seem to be shifting. Early drafts can be produced quickly and fluently. This can be helpful, especially in the early stages of thinking. But it also changes the rhythm of work. There is less time spent sitting with partial ideas, and more emphasis on shaping and editing material that is already well-formed.

In place-based settings, this has a particular edge. AI-generated summaries can give a sense of coherence around issues that are, in reality, complex and contested. In work on catchments or land use, for example, differences in values, knowledge, and priorities are often central to the task. Some of that knowledge is also held through continuing relationships with a place and with other people, rather

than being fully represented in reports or datasets. A technically competent synthesis can therefore miss something important even when it accurately reflects the material it was given.

This connects with something else that is beginning to surface. In the past, much early-career learning has come from working through messy material, drafting, reworking, and making sense of place-specific information. If more of that early work is now supported by AI, it raises questions about how that judgement develops over time, particularly in place-based work where learning is closely tied to context and to working through real, contested situations. What this will mean for the development of judgement is still unclear.

There are also early signals that expectations are beginning to move at an organisational level. Some funders and commissioners are starting to ask about AI use, or to assume a level of AI-assisted work, even where norms are not yet well defined. In some cases, this is happening ahead of clear guidance, leaving teams to work things out in real time as they navigate both the technology and the issues at hand.

What feels less settled

Alongside these more visible shifts, there are also areas where things feel less clear, particularly as these tools begin to shape how people come into conversations and how issues are framed.

One question is how collective judgement is shaped by what enters the room. If participants arrive having used AI to prepare their inputs, this may broaden the range of ideas. At the same time, AI-assisted preparation may introduce a degree of convergence in how issues are framed. In place-based work, where differences in perspective are often central, it is not yet obvious how this will play out.

There are also questions about what happens to the relational qualities that underpin this kind of work. Trust, timing, and presence are not easily captured in generated text, yet they remain central when people are working through contested issues about land, water, or community futures. It is unclear whether these aspects will become more visible as other parts of the work are supported, or whether they risk being taken for granted.

Another area of uncertainty is how organisational cultures will adapt. Some may develop deliberate approaches to AI use, creating space to reflect on how it is used and what it means in context. Others may move more quickly, driven by expectations of efficiency or output, without fully working through the implications. In practice, this variation is already beginning to show in different ways.

More broadly, I find myself wondering about the diversity of perspectives in collaborative processes. In place-based settings, local knowledge, lived experience, and different ways of seeing a situation are often critical. If AI tends to smooth language or align arguments, there is a possibility that these differences become less visible. At the same time, these tools can also help surface new connections or bring in wider perspectives. Both possibilities are plausible, and their relative importance will depend on how these tools are used.

In Aotearoa, these questions can also involve authority over Māori data and knowledge, not simply whether information is represented accurately. The CARE Principles for Indigenous Data Governance, developed with involvement from Te Mana Raraunga, emphasise collective benefit, authority to control, responsibility and ethics in the use of Indigenous data.

There are also wider questions about how expectations of practice may shift over time. If AI-assisted drafting and synthesis become normal, what is seen as competent or thorough work may change. That may bring benefits in some areas, but it may also alter how judgement is recognised and developed. At this stage, these changes are uneven and still forming.

What organisations and teams are starting to work through

In response to these shifts, there is a growing recognition that organisations and agencies need to become more deliberate about how AI is used in practice. In some areas, guidance is beginning to take shape, but much of this is still evolving.

These are not settled approaches, but they point to areas that groups are already working through, where some early indications of useful practice are beginning to take shape:

- Making use visible. Being clear about when and how AI has contributed to drafts, analysis, or synthesis, particularly in shared documents or decision-making processes.
- Keeping responsibility for judgement with those involved. AI can support thinking, but responsibility for interpretation, evidence, and decisions remains with the people and teams involved.
- Checking and grounding outputs. Treating AI-generated material as a starting point that needs to be tested against context, evidence, and local knowledge, rather than accepted at face value.
- Being selective about where it is used. Recognising that some aspects of work, particularly those involving sensitive relationships, local knowledge, or contested issues, may require more care or different approaches.
- Supporting staff and shared learning. Creating space for teams to experiment, reflect, and develop shared understanding about how these tools are used in their particular context.

In place-based settings, these are not just technical choices. They shape how conversations unfold, whose knowledge is visible, and how decisions are made. For many organisations, the challenge is less about having the right policy in place, and more about supporting ongoing discussion and shared judgement as these practices evolve.

What seems worth holding onto

In the midst of these shifts, some aspects of practice still feel steady, particularly where judgement, relationships, and collective sense-making remain central.

- **Judgement still needs to be exercised.** The presence of fluent, well-structured text does not remove the need to ask where ideas come from, how they are supported, and what they mean in a particular place. In work on catchments, land use, or community futures, this often involves working with partial evidence, different knowledge systems, and real consequences over time.
- **Relationships also remain central**. The work of building trust, understanding different perspectives, and staying with difficult conversations is not replaced by faster drafting or synthesis. In many place-based settings, these relationships are what make it possible to work through contested issues and move towards shared action.
- **Collective sense-making continues to matter.** The value in many of the processes I'm involved in lies not just in the outputs, but in the conversations that shape them. It is through these interactions that assumptions are surfaced, differences are worked through, and decisions begin to take form.

These are not simply things that AI cannot do. They are aspects of practice that matter because they are shared, situated, and developed over time, and because they shape what actually happens in the places people care about.

Closing reflection

I've used a futures lens here not to predict what will happen, but to help notice how AI is already entering practice, and where questions are beginning to surface. Different people will be seeing different patterns, depending on the contexts they work in. Some of what I've described may resonate. Other aspects may look quite different elsewhere.

It may be useful to take a moment to consider what is becoming visible in your own practice as AI becomes part of the workplace. What feels like it is changing, what remains uncertain, and what is only just beginning to take shape? Some of those shifts can only be understood by stepping back from practice to the wider organisational and structural conditions within which AI is being developed and used.

PART 4
Stepping back to the wider ethical picture

The earlier sections begin mainly from practice. This final section steps back to consider the wider ethical and structural conditions within which that practice takes place. It looks across practice, organisational and structural levels, then asks what those concerns imply for the design and development of AI itself. The emphasis shifts from how we work with AI to the kinds of systems we are helping to create, and what better development might require.

10. Seeing the wider ethical picture around AI development and use

Debates about AI often sound polarised, and some of that disagreement reflects people focusing on different parts of the system. This essay maps three overlapping levels of concern: the wider systems AI depends on and is beginning to reshape, the organisational choices through which it enters our work, and the practice-level tensions that surface when people use it.

Ethical questions about AI arise at structural, organisational and practice levels. Distinguishing between them helps explain why people can reach very different positions on AI. Often they are worrying about different things, and those concerns need different kinds of response.

Earlier essays in this collection have focused mainly on professional and collaborative practice, including how organisations learn and respond as AI enters that work. Essay 2 considered AI as a thought partner in professional practice, while Essay 3 looked at what happens when AI enters shared and collaborative work. This essay steps back from those particular uses to look at the wider conditions within which AI is developed, introduced and used.

The structural level

The first set of concerns relates to the wider systems within which AI is developed and used, and the ways AI may in turn reshape them. These are not mainly questions about individual behaviour. They concern infrastructure, labour, knowledge and power.

AI as environmental and material infrastructure

AI is often discussed as though it were abstract software. In practice, it depends on data centres, energy supply, water use, mineral extraction and global supply chains. Recent analyses have highlighted pressures on electricity grids and freshwater supplies in some regions. Hardware and model efficiency continue to improve, but overall demand is also growing.

Not all AI systems operate at the scale of large foundation models. Some run locally or use smaller infrastructures. But the generative systems shaping public debate and organisational practice depend heavily on substantial centralised computing resources. Ethical questions therefore concern what resources are used, where the costs fall and what infrastructure societies choose to support.

AI as appropriation of labour and knowledge

A second set of concerns focuses on what AI systems draw upon. Training practices range from fully licensed datasets to highly contested forms of large-scale scraping. This variation fuels much of the debate. For many critics, it raises questions about the appropriation of labour, knowledge and creative

work. Labour here means more than the creative and intellectual work drawn into training data. It also includes the conditions of the work involved in building and maintaining these systems, where critics point to outsourced and low-paid data labelling and content moderation.

Ethical attention here centres on authorship, consent, intellectual property and responsibility for interpretation. Whose work is being used? How was it gathered? Did creators, communities or researchers have meaningful agency in the process?

Public debates over scraped web text, artists' work and open-source code have made these concerns more visible. In research and evaluation, related questions arise about how AI is used in analysis and where responsibility for meaning sits. These concerns may also provide good reasons for placing limits on some uses.

AI as reinforcement of existing power structures

A third structural lens treats AI not only as a tool, but also as a form of governance. The most widely used generative systems are developed and controlled by a relatively small number of organisations. Platform decisions shape defaults, and those defaults influence behaviour. Over time, this affects how knowledge is produced, circulated and valued.

The issue is not only what individual users do. It is also who decides how systems are built, what values are embedded in them and whose interests are served. Governance therefore includes formal regulation and collective oversight, but also the quieter ways platforms shape expectations and define normal practice. Other concerns, including bias, discrimination, harmful uses and longer-horizon questions about model alignment and existential risk, are also important, but sit beyond the scope of this essay.

None of these concerns can be dealt with simply by asking individuals to use AI carefully. Questions about ownership, infrastructure and concentrated power require responses at other levels, including regulation and collective oversight. Principles such as fairness, privacy, transparency, human oversight and accountability, including those set out in the UNESCO Recommendation on the Ethics of AI, provide useful orientation. The challenge is to translate them into decisions and arrangements at organisational and institutional levels.

The organisational level

The second level sits between the wider system and the individual user. Most people do not encounter AI as a general technology. They encounter it as something chosen, paid for, approved or quietly expected by the organisation they work for or alongside. Decisions at that level shape what is available, what is rewarded and what can be questioned.

Adoption without a stated purpose

Organisations introduce AI for different reasons: reducing costs, improving access to information, increasing output, speeding up analysis, supporting staff thinking or freeing time for relational and strategic work. These purposes are not equivalent. They lead to different choices about tools, capability, oversight and acceptable risk.

Yet AI is often introduced before its purpose has been made clear. Adoption may follow from procurement, a platform update or a general sense that others are already using it. The technology then begins to shape workflows and expectations without a shared view of what it is meant to improve.

Where this happens, success is easily judged by uptake, speed or volume rather than by any change in the quality of the work. Questions about who benefits, what may be lost and what other conditions are needed can remain unasked.

I would rather start by asking what we are actually trying to improve and what, specifically, we expect AI to contribute. From there come questions about who benefits, what else needs to be in place, and how we will notice if something unexpected is happening. A Theory of Change can help make these assumptions visible and open to challenge.

Making room for judgement

Organisations may expect people to check AI outputs carefully and apply professional judgement, while also asking them to produce more work in less time. They may encourage experimentation without creating enough space to compare experience, examine mistakes or reflect on what is being learned.

This creates a gap between the responsibility placed on individuals and the conditions available for exercising it well.

Training is often framed around how to write prompts, but the harder capabilities lie elsewhere: assessing sources, noticing weak reasoning, understanding model limits, and recognising when relational, cultural or evaluative judgement should not be handed over. Without these capabilities, people may become confident users and inattentive reviewers.

Transparency that records use but not influence

A third concern is disclosure. Policies and approved-tool lists provide useful boundaries. They matter for privacy, security and responsibility. But a statement that AI was used tells us little on its own. It does not say when AI entered the process, what material it was given, what it helped shape, who checked or challenged it, or where responsibility for the final interpretation sits.

This matters in evaluation, research, policy and collaborative planning, where AI-generated material can influence how a situation is framed and what is treated as credible evidence. A disclosure that satisfies a policy without addressing those questions can leave everyone reassured and no one much better informed.

Responding is less about writing a longer policy than being explicit about purpose and revisiting it as experience accumulates. Organisations need to compare experience, examine failures and notice whether AI is improving the work or simply increasing speed and volume.

The practice level

The wider system matters. So do organisational choices about purpose, capability and oversight. Yet when people use AI in writing, research, evaluation, policy or facilitation, other questions appear. These

concerns arise in everyday practice. Over time, they may also feed back into organisational and structural patterns.

Reliability of reasoning

Generative AI systems write in fluent, confident language. They provide summaries, numbers, examples and arguments. Often this is helpful. It can clarify structure or open up new angles.

But the same fluency can make weaknesses harder to spot. These may include:

- references that appear plausible but cannot be traced;
- statistics presented confidently without a clear source;
- arguments that move too quickly from description to conclusion.

These behaviours are often described as hallucinations or fabrication. However, the term 'hallucination' can mislead by implying randomness, rather than reflecting a recurring feature of how these systems generate plausible language.

When AI-generated material enters research, evaluation or policy advice, we need to ask straightforward questions. Can this be checked? Is the evidence visible? Are limits and uncertainties acknowledged?

In many settings, the risk is not dramatic failure but gradual drift. If unsupported claims are accepted because they are well written, professional standards can quietly shift.

What happens to our own thinking

Another issue concerns what happens to us when we use these tools regularly. When drafting and synthesis are readily available, it becomes easier to hand over early thinking. This can save time and help work move forward. But it may also change how we engage with difficult material. We may:

- spend less time wrestling with uncertainty;
- read more quickly and less deeply;
- move from developing an argument to editing generated text.

These subtle shifts can influence how judgement is exercised and how groups deliberate together.

The effects are not confined to individual thinking. When AI-generated drafts, themes or syntheses enter shared work, they can shape what a group notices and what appears to have been settled, giving an early synthesis more authority than it may deserve. They can provide a useful starting point, but may also narrow the space for alternative interpretations or create a convincing record of participation without people having influenced the framing, interpretation or meaning.

These issues are longstanding challenges in evaluation and collaborative practice. AI does not create them, but it can increase their speed, scale and visibility.

If these changes become widespread, they begin to affect organisational expectations and professional cultures. Workloads may increase, expectations of expertise may shift, and activities that created space for reflection may be treated as unnecessary delays.

Seeing the three levels together

Holding the three levels in view makes disagreement easier to understand. Some people focus on environmental impacts, labour and power. Others are concerned with how AI is introduced where they work, and on what terms. Others again focus on reliability, participation and the effect on human judgement.

The three levels are useful distinctions, but in practice they run into one another. Structural conditions affect what organisations can access and on what terms. Organisations then help determine which systems and uses become normal. And what people do every day can reinforce those arrangements, adapt them or sometimes push back against them.

The diagram below brings these relationships together. It shows the three levels as connected rather than separate, with influence moving in both directions. The concerns shown at each level are indicative rather than complete, and several of them cut across more than one level.

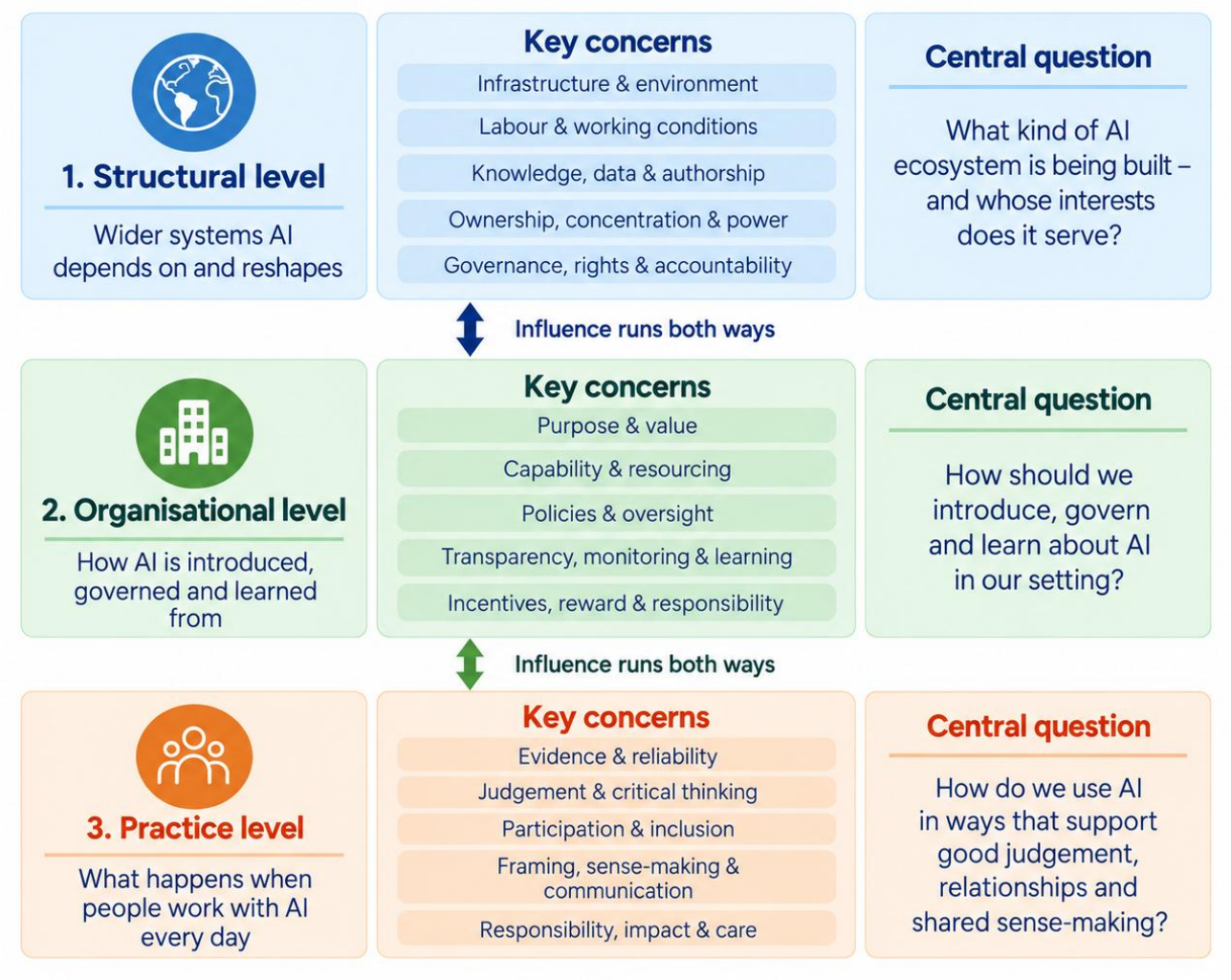

Figure 5: AI ethics across three connected levels. Concerns, choices and practices shape one another.

Influence can move in both directions. Structural conditions shape organisational choices and everyday practice, while organisational decisions and routine uses can reinforce, adapt or sometimes challenge those wider patterns.

A distinction between macro, meso and micro levels is widely used in work on governance, organisations and practice, and similar multi-level approaches are increasingly being explored in discussions of AI ethics. Wang and Blok (2025) for example, distinguish between artifact-level, organisational and broader structural concerns, while emphasising the connections between them. I use a broad version of that distinction here to separate the wider conditions around AI, the organisational choices through which it enters our work, and the everyday practices through which people use it.

No single level provides a sufficient response. Careful individual use does not resolve questions about ownership or environmental impact. Regulation does not give an organisation a purpose for adopting AI. An organisational policy does not ensure that a fluent draft will be read sceptically in a workshop or evaluation.

Responding responsibly therefore requires more than taking a position for or against AI. It involves asking:

- What kind of AI ecosystem are we contributing to?
- How should organisations introduce, govern and learn about AI?
- How do people use AI in ways that support good judgement, relationships and shared sense-making?

Locating my own practice

My own engagement with AI is shaped by concerns at all three levels, although the choices available to me are not the same at each. Individual working habits can help address some practice-level concerns and can contribute to conversations about how AI is introduced in the organisations and groups I work with. They cannot, on their own, resolve wider questions about infrastructure, ownership, labour or concentrated power.

What this has led me to is not a fixed position or a complete response, but a set of working habits.

I use AI purposefully, mainly to support thinking, preparation and early analysis rather than as a substitute for judgement, relationships or craft. I am wary of practices that add speed or volume without improving sense-making.

Source integrity remains important. When AI surfaces articles, statistics or references, I check them before drawing on them.

Where AI is being considered in an organisational setting, I prefer to start with purpose. What are we hoping to improve? What else needs to be in place? Who should be involved, and how will we notice unintended effects?

Questions about AI use need to remain visible in collective settings. This includes naming uncertainty, acknowledging uneven impacts and recognising where responsibility sits.

I have also become more deliberate about examining how these systems behave. I ask for sources, check numbers, test alternative framings and probe the assumptions behind a claim. This is less about catching dramatic failure than understanding recurring patterns of instability and influence.

These practices are, of course, only partial responses. They help keep judgement, evidence and responsibility visible in my own work, and provide a basis for conversations with others. They do not resolve the structural conditions within which AI systems are developed, or replace the need for organisational choices, public governance and collective action.

Before asking how to use AI well, it helps to understand what people are worried about and where those concerns arise. Some things we can address through our own practice. Others require organisational choices, regulation or wider collective action. Being clearer about that distinction seems a better starting point than treating AI ethics as a simple choice between acceptance and rejection.

The next essay takes this wider picture as a starting point and asks what it might require of AI design and development.

11. Designing AI for the wider world: six principles for better development

Essay 10, *Seeing the wider ethical picture around AI development and use*, mapped the wider ethical terrain around AI. This essay asks what those concerns might mean for the way AI is designed, developed and deployed. It draws on six design principles that put purpose, human agency, fairness, planetary limits, wider system effects and accountability at the centre.

That wider view matters because many concerns about AI cannot be resolved simply by asking individuals to use it more carefully. Questions about energy and water use, labour, ownership, concentrated power, changing employment, and effects on education and public information are shaped much further upstream.

Looking upstream brings a different set of questions into view. Rather than beginning with what the technology can do, we can start with what we are trying to achieve, for whom, and within what wider social and ecological setting.

We already know quite a lot about good design

Many of the ideas that might guide better AI development are not new or unique to AI. Participatory development, systems thinking, sustainable design, action research and adaptive management have all had to grapple with a similar problem: what happens when we introduce a technology or other intervention into a setting that is already complex and changing?

Across these traditions there is plenty of accumulated experience to draw on. Technologies have purposes, different people experience their benefits and costs differently, knowledge is distributed, and consequences often unfold in ways that could not have been fully predicted at the outset. The difficulty is not that these ideas are unavailable. It is that they remain far from routine in the development and deployment of powerful technologies. Bringing them into AI design therefore matters not because they are new, but because what is already known about good design is still too easily set aside.

Value-sensitive design provides one example of this connection with AI. Umbrello and van de Poel (2021) argue for bringing values explicitly into AI design and extending attention across the technology's life cycle so that unintended consequences can be identified and responded to.

These traditions also encourage us to question the boundaries around a design problem. It is easy to define the relevant system narrowly around a developer, customer or immediate user. But people live and work in places. We have neighbours and communities. We are linked to people elsewhere through supply chains, labour, information and economic systems. And all of this takes place within ecological systems on which we depend.

So asking whether an AI system works is only part of the question. What is it actually intended to improve? Who gets to decide what improvement means? Who is likely to gain or lose? What else may

change around it? We also need to ask whether the resources involved are justified by the value being created.

Participation is part of this. It is not necessary, or possible, for everyone to participate in every design decision. But good design should ask who needs to be involved, in what, at what stage, and with what influence. This is partly about fairness, but also about knowledge. Designers rarely possess all the knowledge needed to understand how a technology will interact with the situations into which it is introduced.

Six principles for good design

Against that background, I find six principles useful for thinking about better AI design and development. They bring together practical expectations that have emerged across different fields and traditions, and are relevant not only to people designing, developing, deploying and governing AI, but also to those working with other technological innovations.

They are not intended to cover everything that ethical AI requires. Their purpose is to help us ask better questions about purpose, people, resources, wider consequences and responsibility. Participation is not a separate seventh principle because it runs through the others. It can help shape purpose, bring different knowledge and perspectives into design, make consequences more visible and strengthen accountability.

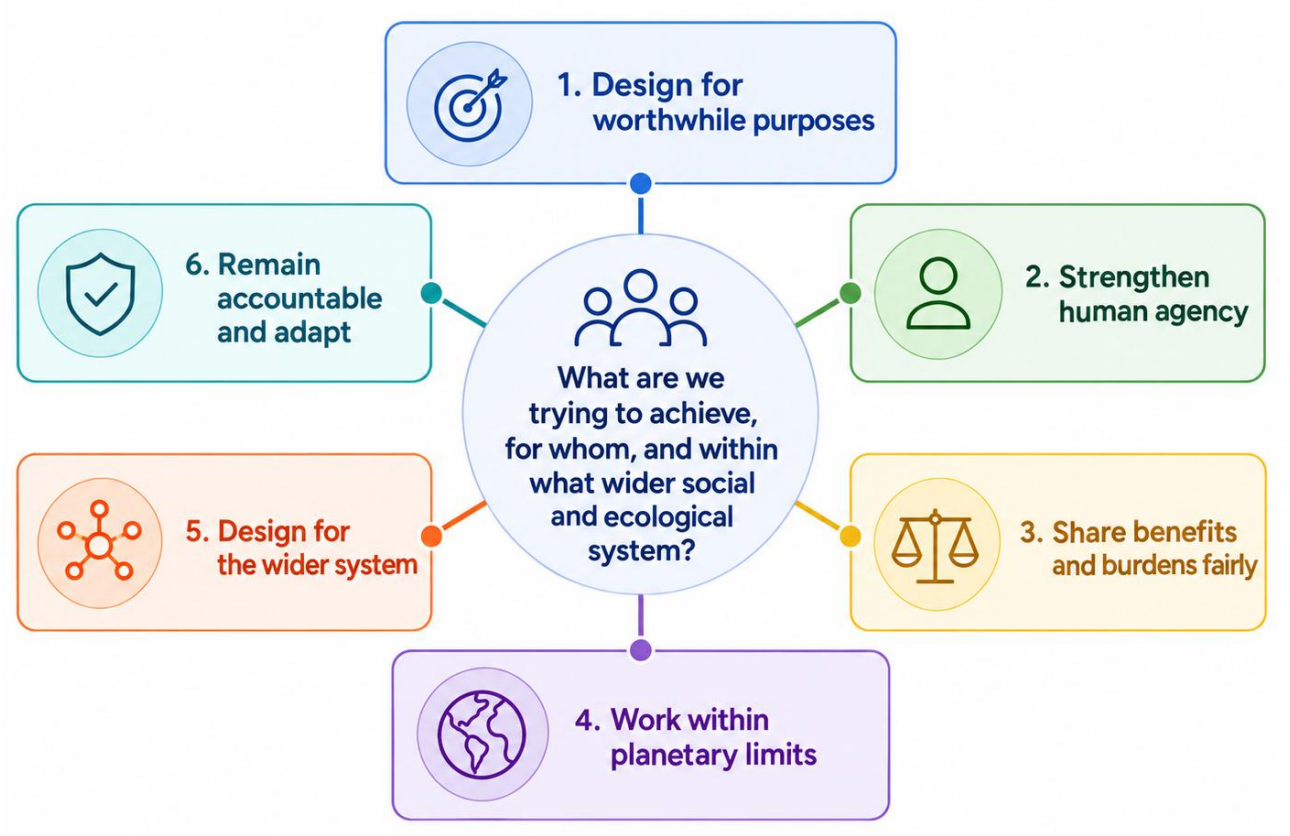

Figure 6: Six principles for the ethical design, development and use of generative AI.

The six principles overlap rather than operating as separate tests. The sections that follow consider each in turn, before returning to how they work together.

1. Design for worthwhile purposes

The starting question should be what we are trying to achieve, rather than what the technology is capable of doing. Technical capability, novelty, market opportunity and productivity can all be relevant, but they are not purposes in themselves. A system may work extremely well at doing something without making the activity particularly worthwhile.

Some fairly basic questions follow. What problem are we trying to address? Who sees it as a problem? Who is expected to benefit? Are there other ways of achieving the same outcome? And what might be displaced or weakened in the process? These questions become particularly important when growing technical capability itself starts to be treated as sufficient reason to automate an activity.

2. Strengthen human agency

Technology can extend human capability. It can help people gain access to information, explore alternatives, communicate more easily and carry out tasks that would otherwise take considerable time. But it can also create dependency, narrow opportunities to learn, reduce professional discretion and quietly shift decisions towards systems that people do not fully understand or control.

Human agency is not only individual. AI-supported systems should also preserve the capacity of teams and organisations to question, deliberate and exercise judgement together.

So better design should aim to strengthen people's capacity to understand, decide, learn and act. Keeping a human formally "in the loop" is not enough if that person's role has been reduced to accepting recommendations they have little capacity or time to question. The more useful question is whether a system supports human judgement and capability over time, or gradually substitutes for them.

3. Share benefits and burdens fairly

AI systems create benefits, costs and risks, and these are rarely shared evenly. Organisations receiving productivity gains may not be the same people whose jobs are changed or lost. People whose creative or intellectual work contributes to AI systems may have little influence over how that work is used. Environmental or infrastructure costs may fall on communities far removed from the users receiving the benefits.

So fairness involves more than checking whether an algorithm treats individual users consistently. It also means asking who gains, who pays, who carries the risks and who has a meaningful say. Those questions need to be considered across the development chain, including data, labour, infrastructure, ownership and deployment. Ownership and control matter here because those able to set the terms of development and deployment may also be better placed to capture the benefits and shift some of the costs elsewhere.

Participation is especially important where those making design choices are not the people most likely to experience their consequences.

4. Work within planetary limits

The material demands of AI were part of the wider ethical picture considered in Essay 10. Here the design question is what follows from recognising them. Improvements in the efficiency of AI systems matter, but so does the overall scale at which they are being developed and used.

Environmental costs need to be part of the design brief rather than treated as something outside it. This does not mean expecting technologies to have no environmental footprint. It does mean making resource use visible, asking where it can be reduced, and considering whether it is proportionate to the value being created. This brings purpose back into the discussion: what kinds of value justify resource-intensive uses of AI, and who should be involved in making that judgement?

5. Design for the wider system

Technologies do not simply enter an existing setting and leave everything else unchanged. People and organisations adapt around them. Practices change, expectations shift, new dependencies develop and activities that were previously valuable may become less visible.

This makes it important to look beyond the immediate user and intended function. How might widespread use affect education, professional practice, employment, communities, information systems or democratic institutions? What behaviours might the system encourage? What might it displace? And what happens if millions of people and organisations respond to the same incentives?

Looking at the wider system will not tell us everything that is going to happen. Complex systems do not work like that. But it can help us notice effects and relationships that are reasonably foreseeable, and make it harder to dismiss them as somebody else's problem. The remaining uncertainty is one reason why continuing learning and adaptation matter.

6. Remain accountable and adapt

No design process will identify every consequence in advance. Complex technologies interact with changing social systems, and effects that seem minor during testing can become significant when systems are widely adopted. Responsible design therefore cannot finish when a product is released.

That means developers and organisations need ways of noticing what happens in practice, learning from unexpected effects, responding to challenge and changing course. People affected by a system need meaningful routes for questioning decisions and seeking redress. Monitoring needs to look beyond technical failure to changes in behaviour, capability, relationships and wider system effects. And accountability needs to remain identifiable. If responsibility is spread so widely that nobody has the authority or obligation to act, it is difficult to call it accountability at all.

These principles work together

The principles overlap. Jobs, for example, raise questions about human agency, fairness and wider system effects. Water and energy use connect planetary limits back to purpose: is this use of resources justified by what the system is helping us achieve? Misinformation brings in the wider information system as well as accountability. Concentrated control over powerful AI systems raises questions about fairness, agency and whose purposes are being served.

The principles can also pull in different directions. A socially valuable use of AI may require significant resources. Wider access may create new risks. Protecting people from harm can conflict with autonomy. Good design will not remove these tensions, but it should make them visible enough to be discussed.

Locating these principles

These principles also reflect the way I have come to think about design and intervention through work in environmental management, participatory practice, systems thinking, evaluation and adaptive management. Across these fields, I have repeatedly run into the limits of approaches that define a problem too narrowly, treat expert knowledge as sufficient, or assume that an intervention will work as intended once it is introduced.

I find the six useful because they bring some of that accumulated learning into the questions now being raised by AI. They offer a set of expectations that may help us think about what better AI development and use could look like.

Concluding thoughts

There is no shortage of principles and standards for responsible AI. What I want these six to do is shift the starting point away from asking what we can make AI do, towards asking what we are trying to achieve, with whom, and what kind of social and ecological system we are helping to create.

These questions are not unique to AI. They arise in many other fields whenever we introduce technologies or other interventions into situations where knowledge is incomplete, people value different things, and consequences unfold over time.

Responsibility does not rest only with technology developers. Governments, organisations, professions, communities and users all help shape how technologies are introduced and what becomes normal practice. But responsibility is not shared equally. Those with greater power to shape AI systems, determine their purposes and influence the conditions under which they are used carry greater responsibility for what follows.

Better individual and organisational use of AI still matters. But it sits within this wider picture. We should also expect better from the technologies themselves, and from those designing and developing them.

That brings us back to the concern running through this collection. As AI becomes part of professional and collaborative work, purpose, judgement, participation, interpretation, learning and responsibility need to remain open to question and influence. Some of that work sits with practitioners and organisations, while some requires action further upstream. Across those levels, the continuing task is to keep what matters visible, discussable and open to judgement as the technology and the practices around it continue to change.

References

Arnstein, S. R. (1969). *A ladder of citizen participation. Journal of the American Institute of Planners*, 35(4), 216–224. https://doi.org/10.1080/01944366908977225

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). *Developing evaluative judgement for a time of generative artificial intelligence. Assessment & Evaluation in Higher Education*, 49(6), 893–905. https://doi.org/10.1080/02602938.2024.2335321

Carroll, S. R., et al. (2020). *The CARE Principles for Indigenous Data Governance. Data Science Journal*, 19, 43. https://doi.org/10.5334/dsj-2020-043

Carter, L. D., & Liu, D. (2025). *How was my performance? Exploring the role of anchoring bias in AI-assisted decision making. International Journal of Information Management*, 82, 102875. https://doi.org/10.1016/j.ijinfomgt.2025.102875

Green, B. (2022). *The flaws of policies requiring human oversight of government algorithms. Computer Law & Security Review*, 45, 105681. https://doi.org/10.1016/j.clsr.2022.105681

Hsu, Y.-C., Huang, T.-H., Verma, H., Mauri, A., Nourbakhsh, I., & Bozzon, A. (2022). *Empowering local communities using artificial intelligence. Patterns*, 3(3), 100449. https://doi.org/10.1016/j.patter.2022.100449

Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2023). *Operationalising AI ethics: barriers, enablers and next steps. AI & Society*, 38, 411–423. https://doi.org/10.1007/s00146-021-01308-8

Samuel, A. (2026). *Learning with machines: Toward a theory of epistemic co-agency. Computers and Education: Artificial Intelligence*, 10, 100573. https://doi.org/10.1016/j.caeai.2026.100573

Sloane, M., Moss, E., Awomolo, O., & Forlano, L. (2022). *Participation is not a design fix for machine learning.* In *Proceedings of EAAMO '22*. ACM. https://doi.org/10.1145/3551624.3555285

Umbrello, S., & van de Poel, I. (2021). *Mapping value sensitive design onto AI for social good principles. AI and Ethics*, 1, 283–296. https://doi.org/10.1007/s43681-021-00038-3

UNESCO. (2022). *Recommendation on the ethics of artificial intelligence.* https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

Wang, H., & Blok, V. (2025). *Why putting artificial intelligence ethics into practice is not enough: Towards a multi-level framework. Big Data & Society* 12(2). https://doi.org/10.1177/20539517251340620

Weidinger, L., et al. (2023). *Sociotechnical safety evaluation of generative AI systems.* arXiv:2310.11986. https://doi.org/10.48550/arXiv.2310.11986

Weidinger, L., Raji, I. D., Wallach, H., Mitchell, M., Wang, A., Salaudeen, O., Bommasani, R., Ganguli, D., Koyejo, S., & Isaac, W. (2025). *Toward an evaluation science for generative AI systems.* arXiv:2503.05336. https://doi.org/10.48550/arXiv.2503.05336

Related resources

The essays in this collection sit within a wider body of practice material developed through the Learning for Sustainability website. While the essays are mainly reflective, many of the ideas they draw on have been shaped through longer-standing work on systems thinking, facilitation, participation, monitoring, evaluation and learning, Theory of Change, adaptive management and reflective practice. The resources below provide practical guidance, frameworks and related material for readers who want to follow those connections further or apply them in their own work.

Generative AI

The main Learning for Sustainability hub for material on AI, professional practice, ethics and organisational use.

Essays on AI and professional practice

The online versions of the essays brought together and revised for this collection.

Monitoring, evaluation and learning

Related material on learning, adaptation, indicators and evaluative judgement.

Theory of Change

Guidance on clarifying purpose, making assumptions visible and linking intended changes with evidence, learning and adaptation.

Systems thinking and complexity-aware thinking

Resources for understanding relationships, complexity, framing and intervention in changing systems.

Reflective and reflexive practice

Resources for supporting reflection, questioning assumptions and learning from experience as professional practice changes.

Supporting change in multi-actor settings

Resources on participation, shared learning, collaborative processes and working with different perspectives.

Climate adaptation / place-based practice

Material connecting adaptive management, learning and long-term decision-making in environmental and place-based settings.

Human ethics in research and evaluation

Practical guidance for independent and collaborative research and evaluation, with particular attention to consent, relationships, power, care and ethical judgement as work unfolds. Includes protocols and guidance for collaborative peer review.

About this collection

Working well in the age of AI brings together essays I wrote between early 2025 and September 2026, drawing on material from Learning for Sustainability. They reflect an evolving inquiry into what happens as generative AI becomes part of professional and collaborative practice, and how those changes connect with organisational choices and wider structural concerns.

The essays are also available individually on the Learning for Sustainability website. Bringing them together here makes the connections between them more visible, while the resources listed on the previous page lead into the wider body of practice material they draw on.

About the author

Will Allen is an independent systems practitioner, evaluator and facilitator based in Ōtautahi Christchurch, Aotearoa New Zealand. For more than 30 years his work has focused on collaborative approaches to environmental management and sustainability, including systems thinking, facilitation, monitoring and evaluation, social learning and adaptive management.

He developed and curates the Learning for Sustainability website, which brings together practical resources for people working with complex environmental, social and organisational change.

A note on AI use

I used generative AI tools during the development of these essays as thought partners for exploring ideas, reviewing structure and language, and assisting with some graphics. I also used them iteratively to develop and test arguments, while determining the overall framing and deciding what to accept, reject and substantially revise. The arguments, interpretations, source checking, editorial decisions and final responsibility for the collection remain mine.

willallen.nz@gmail.com
learningforsustainability.net

learning for sustainability
supporting engagement, co-design and adaptation