

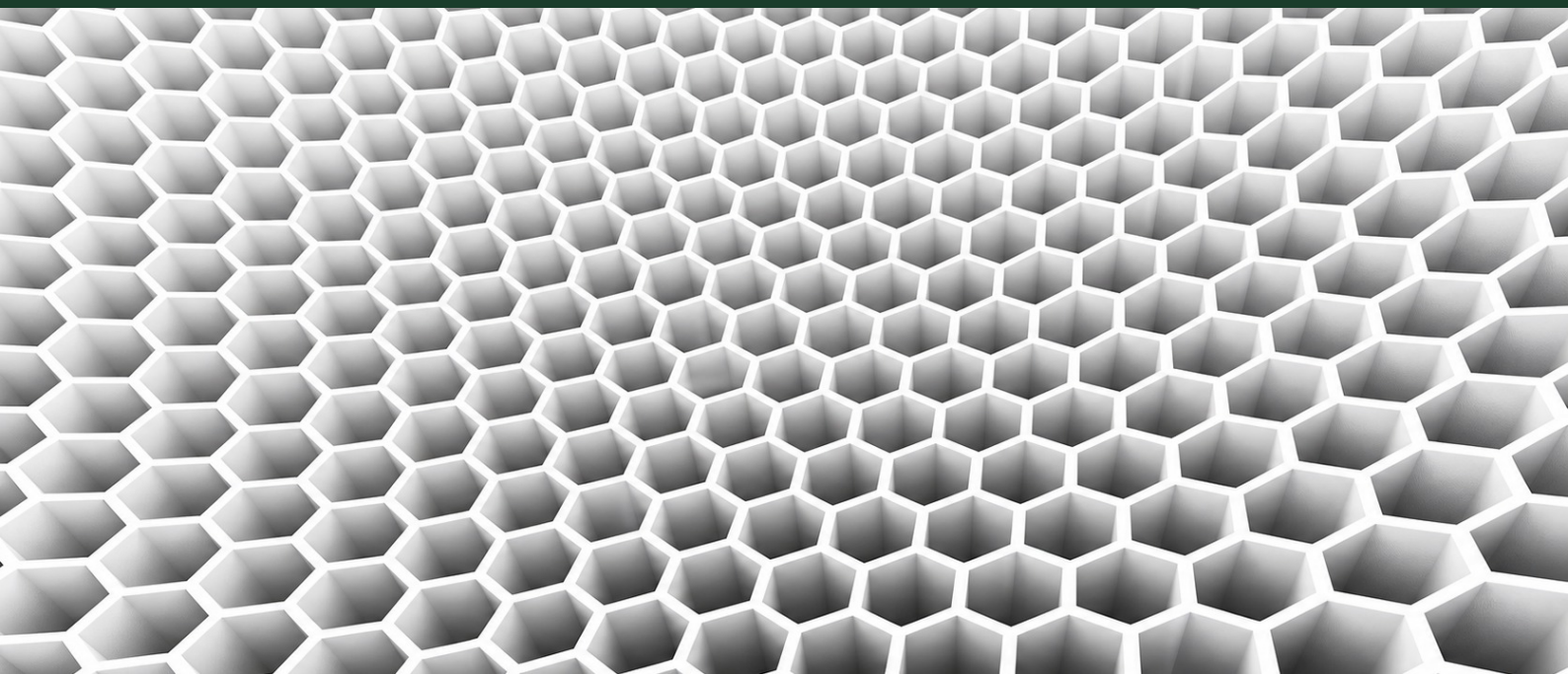
BEYOND PROMPT TRAINING

The AI Skill Shift Is Not About Prompting — It Is About Reclaiming Judgement

What generative AI reveals about programme work, professional learning and the limits of tool-led change

Giovanni Leonardi

January 2024



The AI Skill Shift Is Not About Prompting — It Is About Reclaiming Judgement

Programme · No. 222

Originally written January 2024 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

*“Our visible artefacts are only the residue of judgement;
they are not the judgement itself.”*

1. The training room and the real work

At nine o'clock, a group of programme professionals are shown how to ask a generative AI system for a risk register. By lunch, they can vary the instruction, specify a tone and request a table. The output is quick, plausible and neatly categorised. It also misses the risk that matters: a supplier dependency which was never written down because everyone in the programme assumes everyone else knows it.

The lesson is uncomfortable. The new tool has exposed an old truth about programme work. Our visible artefacts are only the residue of judgement; they are not the judgement itself.

This matters in January 2024 because organisations are beginning to translate the excitement around large language models into capability programmes. Much of that effort is framed as a skills gap: teach people to prompt, teach them to check outputs, teach them the approved tool. Useful, certainly. Sufficient, no.

The deeper skill shift is not from writing documents manually to writing instructions for a machine. It is from producing programme artefacts to designing, testing and governing the decisions those artefacts are meant to support.

That shift reveals something important about how organisations change. They prefer to train individuals in a new technique because technique is visible, purchasable and countable. They avoid changing authority, workflow and professional expectations because those changes disturb the operating model. The result is familiar: transformation intent is ambitious, while transformation reality is reduced to a course catalogue.

2. The seduction of the visible skill

Prompting has become the emblem of AI fluency because it produces an immediate demonstration. A weak request returns a generic answer; a stronger request returns something better. The improvement feels teachable. It can be shown in an hour and measured with a completion badge.

But programme professionals do not create value because they can phrase a request elegantly. They create value because they know which question deserves an answer, what evidence is missing, whose interests are concealed, which commitment is reversible and when apparent agreement is merely fatigue.

A model can draft a dependency map from supplied material. It cannot notice the dependency that the organisation has chosen not to record unless someone provides the clue. It can summarise a steering discussion. It cannot decide whether the silence of one executive means consent, uncertainty or political withdrawal. It can produce options. It does not carry the consequences of choosing among them.

Prompting improves the interface. Programme judgement determines whether the interface is pointed at the right work.

The strongest case for prioritising technical fluency is practical. People cannot exercise judgement through tools they do not understand. Without basic experience of context limits, variable outputs, source grounding and confident error, programme leaders will either trust too much or reject too quickly. This is correct. Technical fluency is entry-level competence. The mistake is to treat the entrance as the destination.

3. What the machine makes visible

New tools often reveal the hidden structure of a profession. When routine production becomes easier, the work that remains is the work we had bundled inside production without naming it.

A weekly programme report offers a simple example. On the surface, it is a document assembled from workstream updates. Underneath, it contains several kinds of judgement:

- deciding which changes are material;
- distinguishing delay from emerging failure;
- reconciling incompatible accounts;
- identifying where a decision is being deferred;
- framing uncertainty without creating false alarm;
- directing attention toward the next intervention.

Generative AI can accelerate the surface assembly. If the professional role is defined mainly by assembling text and slides, the role will look threatened. If the role is defined by the quality of attention and intervention, the tool may remove low-value friction and expose how little time has been reserved for the real task.

The uncomfortable part is that some organisations have rewarded the artefact more reliably than the judgement. Templates are inspected; decisions are not. Reports are due on Thursday; the quality of challenge has no calendar entry. A polished pack is easier to govern than a difficult conversation. AI does not create this distortion. It makes it harder to ignore.

When the cost of producing the artefact falls, the organisation discovers whether it ever valued the judgement behind it.

4. The skills beneath the skill shift

The capabilities now needed are not exotic additions to programme management. They are neglected parts of the discipline, sharpened by a system that can generate convincing material at speed.

4.1 Problem framing

The first skill is deciding what the system should and should not solve. A request to “identify programme risks” is too broad. A useful frame names the decision, the time horizon, the evidence boundary and the consequences of omission.

Problem framing also means refusing the use case that automates activity without improving an outcome. If a team spends four hours producing a report nobody uses to decide, reducing that work to forty minutes is an efficiency gain inside a governance failure.

4.2 Evidence design

Programme professionals must become more exact about provenance. Which sources are authoritative? Which are current? What remains disputed? What was inferred rather than stated? A fluent answer that collapses those distinctions is dangerous precisely because it is easy to read.

Evidence design is not data science in miniature. It is the professional habit of making the chain from source to claim inspectable.

4.3 Work decomposition

The practical question is rarely whether AI can “do programme management.” It is which parts of a workflow involve retrieval, transformation, comparison, judgement, negotiation, authorisation and action.

Work element	Suitable contribution	Human responsibility
Retrieval	Find relevant material across approved sources	Define authority and completeness
Transformation	Summarise, classify, reformat or draft	Test whether meaning has changed
Comparison	Surface differences and patterns	Decide materiality and consequence

Work element	Suitable contribution	Human responsibility
Judgement	Generate options and challenge assumptions	Own the decision and its trade-offs
Communication	Draft for a defined audience	Carry intent, timing and relationship
Action	Prepare or execute bounded steps	Set permission, limits and recovery

Decomposition protects against two equal errors: assuming the whole profession can be automated, and assuming none of it can.

4.4 Evaluation

Programme professionals are accustomed to acceptance criteria for systems but often less disciplined about criteria for generated work. “Looks good” is not evaluation. A useful test set includes routine cases, ambiguous cases, missing evidence and known failure patterns.

Evaluation must be tied to the decision. A summary can be grammatically excellent and operationally useless if it removes the only qualification a sponsor needed to see.

4.5 Decision governance

The professional must know where assistance ends and authority begins. If a system recommends a change, who may accept it? If it updates a record, what evidence is preserved? If it is wrong, how is the action reversed?

This is not a new compliance layer wrapped around innovation. It is the ordinary work of decision rights, controls and accountability applied to a new form of contribution.

4.6 Learning in the work

Formal training cannot keep pace with a field changing this quickly. Capability will develop through supervised practice: use a bounded workflow, compare output with experienced judgement, record failure, refine the design and widen authority only when evidence supports it.

The unit of learning should therefore be the workflow, not the individual prompt. The organisation learns when an improved practice becomes repeatable across people and cases.

5. A composite programme office learns the difference

Consider a programme office of twelve people supporting a large operational change. Its first AI pilot focuses on monthly reporting. Staff are given prompt training and encouraged to use an approved assistant to summarise workstream updates.

After six weeks, drafting time falls by 38 per cent. Adoption is high. The pilot is declared successful.

Yet board members report no improvement in decision quality. Three material dependencies appear late. One workstream's optimistic wording is reproduced almost unchanged. Analysts spend less time drafting but more time checking whether summaries have removed caveats. The programme has accelerated report production without improving programme sensing.

The office changes the experiment.

Instead of asking each analyst to become a better prompter, it redesigns the reporting workflow. Workstream submissions separate confirmed fact, forecast, assumption and request for decision. Source links travel with each claim. The assistant compares changes across reporting periods rather than merely summarising the latest text. Analysts review exceptions and contradictions. The programme manager owns the final judgement about materiality.

The result is less dramatic on the original metric: drafting time falls by 24 per cent, not 38. But the board pack shrinks from 46 pages to 29. The number of issues requiring board attention is reduced from 31 to nine. Four unresolved dependencies are exposed one cycle earlier. The time saved by analysts is assigned to evidence checking and decision preparation.

What changed was not primarily prompt quality. The office developed a clearer model of its own work.

6. Why organisations keep teaching the wrong thing

The persistence of shallow capability programmes is structural.

Training departments need a defined curriculum, while the necessary learning crosses technology, operations, governance and professional practice. Technology teams can explain tool behaviour but may not own the programme decisions affected. Programme functions understand the decisions but may lack permission to redesign systems or access. Risk teams enter late and focus on prohibition because the operating choices have already been made.

There is also a more human force. Teaching a new technique leaves existing status largely intact. Redefining professional value is more threatening. If report production no longer proves diligence, what does? If access to information no longer confers authority, what does? If junior staff can generate an acceptable first draft, how will they learn the tacit judgement previously acquired through drafting?

These are not objections to adoption. They are the real agenda of adoption.

The apprenticeship question is especially important. Routine work has often been both burden and classroom. Removing it can be beneficial, but only if organisations create new ways for less experienced professionals to observe reasoning, compare alternatives and receive feedback. Otherwise efficiency at the bottom of the profession may create fragility at the top several years later.

7. From capability catalogue to professional standard

A credible response begins by defining AI-era programme competence through outcomes rather than tool features.

A capable programme professional should be able to:

- frame a consequential task with explicit boundaries;
- decompose the workflow into assistance, judgement and authority;
- identify authoritative evidence and preserve uncertainty;
- evaluate output against representative and adverse cases;
- design human intervention where consequence demands it;
- explain the chain from source to recommendation;
- measure whether the workflow improves an operating decision;
- recognise when not to automate.

Technical knowledge belongs inside this standard: how context is supplied, why outputs vary, how retrieval can fail, how sensitive information is handled, and how tool permissions alter risk. But technical knowledge serves professional judgement; it does not replace it.

Development should combine short instruction, observed practice and assessed application. Ask professionals to redesign one real workflow, produce an evidence and decision map, test the intervention against difficult cases, and present what changed. The assessment is not whether they obtained a clever answer. It is whether they made the work more reliable.

8. The profession after the paperwork

There is a pessimistic reading of the present moment: programme work will be hollowed out as machines absorb analysis, drafting and coordination. There is an optimistic reading: professionals will be liberated from administration and elevated into strategic leadership.

Both are too easy. Technology does not decide which version arrives. Organisations do, through the work they reward, the authority they distribute and the learning they protect.

If they continue to define programme professionalism through artefact production, automation will reduce roles without improving change. If they use the moment to restore judgement, evidence, challenge and intervention to the centre of the profession, the work may become both smaller and more valuable.

The skill shift is therefore a test of transformation maturity. Can an organisation change the definition of good work, or can it only add a new tool to the old job description?

Programme professionals do need to learn how generative AI behaves. More importantly, they need to learn to see their own work with greater precision. What decision is this activity serving? What evidence makes it defensible? What part requires judgement? Where does authority sit? What will improve if the activity becomes faster?

Those questions are older than the technology. The technology has simply made them impossible to postpone.