EVIDENCE PAPER · FREE EDITION · NOVEMBER 2026

Before the Prompt

Judgment, framing and calibration as human contributions to AI-assisted work

For HR leaders, hiring managers and anyone writing an internal proposal about how people and generative AI share creative work.

> PROJECT PERSPECTIVE
>
> *We begin with a conviction: common sense is part of being human. People bring perception, intention and responsibility to their work. AI develops its capabilities through training on data.*
>
> This is the project's philosophical position, not a finding of the studies cited in this paper.

How to read this paper. Every claim is tagged. **EVIDENCE** means a cited, published source says it. **INTERPRETATION** is our reading of that evidence. **PROPOSED** is a practical method we suggest; it has not been validated in trials.

Human Creative Value in AI — provisional project name. Not peer reviewed.

1 · THE ARGUMENT

Recognising fit is not the same as producing it

A person can see that an image is almost right — that the light is too harsh, that the hand looks wrong, that the tone is a little too eager — without knowing how to fix it in Photoshop. Generative tools now carry much of the production work. That makes a different skill more visible: knowing what the result should be, and recognising when it is there.

INTERPRETATION

Qualitative judgment and manual tool proficiency are separable capabilities. They often travel together, because people historically learned taste through making. But one does not strictly require the other, and hiring processes that screen only for tool names may miss people with strong calibrating judgment.

EVIDENCE

Jacobi and colleagues (2024) found meaningful individual differences in scientists' aesthetic disposition, aesthetic experiences and aesthetic sensitivity within scientific work — a field not usually seen as aesthetic. Aesthetic engagement varies between people, and it appears in work far from the arts.

What this paper does not claim

Our view that common sense is part of being human is a conviction, not something these studies test. The evaluative judgment discussed here varies between people and can be developed with training; we do not claim it is fixed. We do not claim formal experience is irrelevant. And we do not claim AI review is useless: Kranzle and Sharratt (2025) compared GPT models with human experts in evaluating ideas, which shows automated evaluation is a serious research subject in its own right.

2 · FRAMING BEFORE GENERATION

The most creative act may happen before the prompt

EVIDENCE

Doshi and Hauser (2024) found in an online experiment that access to generative AI ideas made individual stories be rated as more creative, yet the stories became more similar to each other. Gains for one person can come with less variety across a group.

EVIDENCE

Maier, Schneider and Feuerriegel (2026, CHI) experimentally compare model-led and human-led interaction in human–AI co-creation. The design question itself — who leads the interaction — is now studied directly. Readers should consult the paper for its specific results.

INTERPRETATION

If the tool tends to pull toward common answers, then the distinctive human contribution moves earlier: deciding what problem is worth solving, what must be different, and who it is for — before anything is generated.

3 · CALIBRATION

Not more — exactly enough

ILLUSTRATION · CONSTRUCTED ANALOGY

A diner says the steak is undercooked. It comes back burnt. The instruction “cook it more” was binary; the diner wanted a warm pink centre. The correction overshot the complaint.

> “Cook it more” is a direction. “A warm pink centre, keep the seasoning” is a target.

INTERPRETATION

Small qualitative corrections to AI output — copy, image, design, performance — can produce large unintended changes. Four patterns are worth keeping apart, because they need different responses:

- **Overcorrection** — the change goes past the intended middle.
- **Regression** — fixing one constraint breaks another that was already right.
- **Drift** — unrequested changes spread beyond the scope of the request.
- **Repetition** — the same state returns after equivalent attempts, with nothing new learned.

We name these as practical categories, not as proven technical faults with a single cause.

4 · PRESERVE WHAT IS ALREADY RIGHT

Every correction needs a keep-list

PROPOSED

When asking for a revision, state three things: what to change, what to keep exactly as it is, and how you will judge whether the result is right. The keep-list is often the part people forget — and the part that prevents regression and drift.

PROPOSED

Let the person who holds the relevant standard judge the result: an editor, creative director, clinician, sound director, teacher or designer — not necessarily the engineer operating the tool. A second AI reviewer can catch errors when it has independent criteria, but adding another agent does not by itself supply a qualified standard.

EVIDENCE

NIST's Generative AI Profile (NIST AI 600-1, 2024) frames generative AI risk management around governance, mapping, measurement and management, including human oversight and documented review — a useful anchor for internal policy.

5 · FOR HIRING AND INTERNAL PROPOSALS

Look for calibration, not only tool names

EVIDENCE

Sackett, Zhang, Berry and Lievens (2023) revisit the design of selection systems after revised validity estimates for widely used predictors. Among their themes: structured approaches deserve more weight, and the design of a selection system should follow from evidence about what predicts performance.

PROPOSED — A CALIBRATION AUDITION (NOT VALIDATED)

1. Show a candidate a near-miss result and an overcorrected one.
2. Ask what they would preserve, what they would change, and how they would know it is right.
3. Have independent, qualified evaluators score answers against criteria agreed in advance.
4. Pilot the exercise for reliability and fairness before using it in decisions.
5. Keep the hiring decision with accountable people.

This exercise has not been tested. It is offered as a starting point for an internal pilot, not as a validated assessment.

Limitations

This is a short editorial synthesis, not a systematic review. Cited studies use specific samples and tasks; results may not transfer to your organisation. Cost figures for correction loops (rework, review time, human hours) are not measured here and should not be assumed to be zero.

REFERENCES

Sources

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*. https://doi.org/10.1126/sciadv.adn5290

Jacobi, Varga, Jessani & Vaidyanathan (2024). Individual differences in scientists' aesthetic disposition, aesthetic experiences, and aesthetic sensitivity in scientific work. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2023.1197870

Kranzle & Sharratt (2025). Evaluating creative output with generative artificial intelligence: Comparing GPT models and human experts in idea evaluation. *Creativity and Innovation Management*. https://doi.org/10.1111/caim.70007

Maier, Schneider & Feuerriegel (2026). Partnering with generative AI: Experimental evaluation of model-led and human-led interaction in human-AI co-creation. *Proceedings of CHI 2026*. https://doi.org/10.1145/3772318.3791185

National Institute of Standards and Technology (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). https://doi.org/10.6028/NIST.AI.600-1

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2023). Revisiting the design of selection systems in light of new findings regarding the validity of widely used predictors. *Industrial and Organizational Psychology*. https://doi.org/10.1017/iop.2023.24