01 / THE DISTINCTION

The Eye Without the Tool

Recognising qualitative fit is not the same as manually producing it.

A person may recognise that a composition, explanation or performance misses its intent without knowing Photoshop, editing software or code. That person may help calibrate direction while a technical practitioner executes the change. This is an interpretation about complementary roles—not a validated talent category or a claim that software skill is irrelevant.

The business question

Does your selection process identify people who can explain what to preserve, what to change and how to verify the result—or mainly people who can list the tools they operate? For HR teams and internal proposal authors, this paper offers an evidence-led starting point, not a ready-made hiring test.

Three evidence levels

Research: published studies and risk guidance, cited [1–6].
Interpretation: a role distinction developed from that research.
Proposal: a locally evaluated audition and acceptance process. No new hiring instrument or business outcome has been validated here.

Our perspective: common sense is innate, not manufactured by an exercise. This philosophical position is not established by the cited research. Evaluative talent may exist apart from software proficiency; recognition remains fallible.

02 / WHAT RESEARCH SUPPORTS

Sensitivity is real.
Transfer is not automatic.

Aesthetic engagement beyond the arts

Jacobi and colleagues [1] studied 3,092 biologists and physicists across four countries. Their survey found associations between personality traits and aesthetic disposition, experiences and sensitivity in scientific work. This supports variation in aesthetic engagement outside professional arts practice. It does not establish causal mechanisms, a fixed biological essence, or the validity of an HR test for visual taste.

Tacit recognition needs an observable account

Our interpretation: someone may notice a near-miss before fully explaining it. Treat that first response as a hypothesis. Ask for a concrete comparison, the intended audience or task, and what evidence would change their mind. The reviewed sources do not directly validate a general “eye without tool” construct or prove that recognition and execution are independent abilities.

Selection evidence, not résumé rejection

Sackett and colleagues [2] reconsider how selection-system design should respond to revised validity estimates, including variability and validity–diversity tradeoffs. Their work is a reason to examine assessment evidence carefully, not to declare all experience irrelevant or to replace a résumé with an untested aesthetic audition.

Implication for an internal proposal

Define the job-relevant judgment first. Retain relevant experience and technical competence where the role requires them. Add an accessible performance exercise only if it improves the evidence for that role. Do not use personality findings as a shortcut to screen applicants.

03 / AI AS A PARTNER

Better output is not the only outcome.

Creative assistance can help—and narrow

Doshi and Hauser [3] found in an online short-story experiment that access to generative-AI ideas improved evaluated creativity, writing and enjoyment, particularly for less creative writers, while AI-assisted stories became more similar to one another. This is evidence about that task and intervention, not a universal finding about all creative work or every model.

AI evaluation is possible, conditionally

Kranzle and Sharratt [4] report moderate alignment between GPT-4 and expert idea evaluations when using a base framework and spectrum-based few-shot prompting. Their findings make design choices and references important. They do not justify saying AI cannot review AI, or that model ratings always match expert judgment across domains.

Interaction design affects ownership

Maier, Schneider and Feuerriegel [5] report an experiment comparing model-led rewriting with reflective human-led interaction, and validation in another context. Both approaches improved idea quality; the human-led modes preserved diversity and perceived ownership better in the reported settings. Transfer to a particular organisation, profession or hiring process remains to be checked.

Practical interpretation

Use AI to surface alternatives and compare output with explicit references. Retain a qualified acceptance standard and accountable human review. A chain of agents without that standard can repeat or amplify a mistake; independent tests and reviewers can also help. Neither outcome is inevitable.

04 / CONSTRUCTED ILLUSTRATION

Undercooked.
Returned burnt.

A diner says the steak is undercooked. It returns burnt. The requested middle was never understood. This is a constructed analogy, not a reported restaurant incident or experimental evidence.

The chef’s calibrated culinary expertise defines the target; the ordering-system engineer can support execution and communication. Job titles alone do not guarantee expertise. The question is who can interpret the intended quality and verify an appropriate correction.

Four different patterns

Overcorrection: a requested adjustment overshoots the intended target.
Regression: a correction damages something that previously worked.
Drift: the output moves away from the agreed intent or scope.
Repetition: another attempt leaves the observed failed state unchanged.

These patterns can coexist, but they are not one proven technical fault. A narrow correction may require nuanced calibration—not a binary reversal or another cosmetic variation of the same instruction.

Keep reported evidence separate

The site’s reported homepage drift (11 November), separate comparison-illustration anomaly (12 November), unapproved roughly 60-second film review (12 November; director’s subjective 5.2/10), and repeated connector state (12 November; five attempts self-reported) remain distinct. They are contextual observations, not independently verified trials. The film’s final acceptance is unverified; some assets may be reusable. Cause, human hours and financial effects are not established.

Count direct tool use, attention and specialist review separately; never convert missing cost data to zero or count the same hour twice. No saving is demonstrated here.

05 / PROPOSED HR PILOT

Show the near-miss. Ask what must survive.

A job-relevant audition—not a validated instrument

Define the criterion. A qualified domain reviewer and HR lead agree the task, intended audience, essential constraints and what “acceptable” means. Keep technical operation separate unless it is genuinely part of the job.

Use paired material. Offer a brief, an initial near-miss and an overcorrected alternative. Ask the candidate what they notice, what to preserve, what to change and how they would verify it. Let them explain verbally or in another accessible format; do not require proprietary software for a recognition-focused role.

Calibrate independent review. Qualified evaluators first assess the same material independently, then compare rationales. Score job-relevant specificity, preservation of constraints, use of evidence and willingness to revise—not taste conformity, confidence or personal similarity to the panel.

Pilot before selection use. Check agreement between raters, clarity, candidate burden, accessibility and possible group differences. Use relevant performance criteria to investigate whether the exercise adds useful information beyond existing selection evidence. Establish a review and appeal route; keep consequential decisions human accountable.

Internal business-case discipline

Describe the decision this pilot could improve, responsible reviewers and a bounded evaluation period. Record tool spending, recruiter attention and specialist review time as separate categories. Define success criteria before use; collect appropriate evidence with consent and privacy safeguards. Do not promise ROI, fairness or predictive validity before evaluation.

NIST’s Generative AI Profile [6] is risk-management guidance, not a validation of this pilot. Use it to support documented responsibilities, evaluation and monitoring.

06 / REFERENCES & LIMITS

Sources you can inspect

[1] Jacobi, C., Varga, P. J., Jessani, Z. & Vaidyanathan, B. (2024). Individual differences in scientists' aesthetic disposition, aesthetic experiences, and aesthetic sensitivity in scientific work. Frontiers in Psychology, 14, 1197870.
https://doi.org/10.3389/fpsyg.2023.1197870

[2] 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, 16, 283–300.
https://doi.org/10.1017/iop.2023.24

[3] Doshi, A. R. & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10, eadn5290.
https://doi.org/10.1126/sciadv.adn5290

[4] Kranzle, T. & Sharratt, K. (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

[5] Maier, S., Schneider, M. & Feuerriegel, S. (2026). Partnering with Generative AI: Experimental Evaluation of Model-Led and Human-Led Interaction in Human-AI Co-Creation. CHI 2026.
https://doi.org/10.1145/3772318.3791185

[6] NIST (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1. This is the GenAI companion profile, not the base AI RMF 1.0.
https://doi.org/10.6028/NIST.AI.600-1

Evidence and review limits

A focused narrative synthesis, not a systematic review, legal advice or a validated hiring assessment. Publication metadata and accessible abstracts/records were checked for this edition; effect sizes and full methodological appraisal are not supplied. No prospective evaluation, independent case audit or measured business benefit is claimed. Recommendations on page 5 are proposals.