Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Credential farming in the age of generative AI: relocating trust from outputs to communities

Gift Osei Agyemang
University of Energy and Natural Resources
Rawad AL-Louzi
Al-Balqa Applied University
Melat Mezgebu Bawoke
melat.mezgebu@aau.edu.et

Addis Ababa University
Connor Bender
Hackensack Meridian School of Medicine
Leo Anthony Celi
Massachusetts Institute of Technology
Nishita Chavan
East Brunswick High School
Rahul Gorijavolu
Johns Hopkins University
Diego Halac
Sanatorio Anchorena del Callao
Nikhil Jaiswal
Massachusetts Institute of Technology
Sofiya Mikolutskaya
Anthony Nguyen
Troy High School
Raul Ignacio Rojas Sillery
University of Florida
Sparsh Roy
Hopewell Valley Central High School
Dharaa Suresha
Dwight Global Online School
Alexa Xiang
The Ohio State University
Boya Zhang

Massachusetts Institute of Technology

comment

Keywords:

Posted Date: October 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10944586/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: Competing interest reported. RG serves as a course lead and paid consultant for the Collaborative Institutional Training Initiative (CITI Program) in the development of an online course on artificial intelligence for healthcare providers. The remaining authors declare no competing interests.

Abstract

Publications, affiliations and authorship positions began as proxies for qualities that cannot be observed directly: judgement, understanding and the capacity to contribute to shared inquiry. Generative artificial intelligence has weakened the inference those proxies depend on. Polished research outputs can now be produced by people who do not understand them, yet admissions committees and hiring panels still read an output as evidence of the competence that producing it once required. The result is faster credential farming, the extraction of the social trust held in credentials without contribution to the enterprise that gives them meaning. Better detection of machine-generated text will not repair this. What has to change is where epistemic trust sits. Work in the sociology of credentials, in social epistemology and on situated learning points to where it should move: scientific formation, how communities shape people and teach them what to value, carries the evidential weight that credentials only summarised. Three commitments follow. Select members for depth rather than legibility. Build small relational structures in which judgement can be formed and witnessed. Refuse to convert those structures into new metrics. The argument matters most for researchers in low-resource settings, whose learning happens in informal spaces that credentialing systems do not see, and who are asked to make themselves legible in a vocabulary nobody taught them.

Introduction

A secondary-school student can now co-author a polished preprint, list a laboratory affiliation and present to an admissions committee as a scientific contributor, without having learned much in the process. Little in the resulting portfolio separates that student from a peer who spent two years earning the same lines. The problem is evidential. The documents that institutions use to infer competence no longer support that inference reliably.

Credentials began as proxies for qualities nobody can observe directly (Spence, 1973), and generative artificial intelligence has loosened the link between producing a research artefact and possessing them. An arms race of detection will not restore it. Trust has to move off outputs, which can be manufactured, and onto formation inside research communities, which is harder to fake. Selection and mentorship should be rebuilt on that basis.

This argument comes out of discussions within MIT Critical Data, a global open-science community built around a widely used critical care database (Johnson et al., 2016). It does not speak for any institution. We write from inside the tension we are describing: communities that benefit from the symbolic power of elite affiliations attract, by that same power, people who want the symbol more than the work.

Credentials were always proxies

A credential is compressed testimony. A doctorate certified that its holder had survived years of supervised apprenticeship; a first-authored paper certified that its author did the intellectual work the byline implied. Collins (1979) argued that credentials did something else at the same time: they rationed

access to status and worked as currency in a stratification system only loosely coupled to competence. The two functions coexisted for most of the twentieth century, because producing the artefact required the competence.

What happens to proxies under pressure is well documented. High-stakes quantitative indicators corrupt the processes they are meant to track (Campbell, 1979), and, in Strathern's (1997) formulation of Goodhart's (1975) law, a measure that becomes a target ceases to be a good measure. Academic science has supplied three decades of illustration. The h-index and the journal impact factor became units of professional advancement, and researchers learned to optimise the unit rather than the understanding it once indicated (Muller, 2018). The research-assessment community has conceded the point in principle: both the Leiden Manifesto and the San Francisco Declaration on Research Assessment tell evaluators to stop outsourcing judgement to metrics (DORA, 2013; Hicks et al., 2015). The metrics persist anyway, because they still predict who gets scarce positions even as their epistemic meaning drains away.

The motivational residue of all that optimising matters here. Extrinsic rewards can crowd out intrinsic interest (Deci and Ryan, 1985), and intrinsic and prosocial motives are associated with more creative and more useful work (Grant and Berry, 2011). Merton (1973) counted communism, the shared ownership of findings, among the norms of science; an enterprise built on sharing depends on members who are motivated to contribute rather than to extract. Credential farming, our term for taking the social trust held in credentials without contributing to the enterprise that gives them meaning, is therefore more than a fairness problem. Over time it changes what science gets done.

Figure 1 sets the two dynamics side by side. The difference is structural. Credential farming closes into a loop, in which the weakened signal becomes the reason to produce more output, so the cycle accelerates as its evidential value falls. Formation does not close. It accumulates, and each stage leaves behind someone who can vouch for what happened.

What generative AI changed

Every credential system runs on one inference: the artefact exists, so its author has the competence that producing it requires. Large language models have weakened that inference. Polished prose, literature syntheses, working code and plausible analyses are now available to people who understand little of it (Kasneci et al., 2023). Much of the education literature treats this as an integrity crisis to be met with detection (e.g. Cotton et al., 2024). That framing is too small. Detection tools are unreliable and biased against writers whose first language is not English (Liang et al., 2023; Weber-Wulff et al., 2023), and they defend a boundary that has stopped marking much. The damage is to validity: the artefact was evidence of competence because of what producing it demanded, and it demands less now (Dawson et al., 2024).

A fast-growing literature makes a related diagnosis. Signalling accounts hold that generative AI compresses the differential cost of producing academic outputs and so erodes their credibility as evidence of competence (Tamir and Nairz-Wirth, 2026). Work on assessment argues that the

assumption linking the quality of a written product to the ability of its producer has quietly collapsed (Inoshita et al., 2026). Interview studies of doctoral students describe scholarly labour shifting from production towards verification (Wu, 2026). These accounts locate the remedy in verification: better checks, oral examination, proof of process. We accept the diagnosis and press the remedy further. Verification still treats the individual artefact as the unit of evidence and asks only that it be checked more carefully. Our claim is that the evidential weight was never in the artefact. It lies in formation, in the sustained and witnessed participation through which a community comes to know what someone can do, and evaluators should look there rather than at outputs. That claim also reaches researchers formed in low-resource settings, whose formation happens in spaces institutions do not see and whom the verification literature has not addressed.

Three qualifications guard against overstating this. AI has not made knowledge worthless. It has made unwitnessed output weaker evidence of judgement and responsibility. Someone who uses a model as an instrument of thinking may learn more and faster, and the difficulty is that their artefacts look much like those of someone who learned nothing. The inference was also fraying well before 2022, as paper mills and honorary bylines diluted it. Generative AI did not invent the split between production and comprehension. It made the split cheap and available to anyone who can write a prompt. Nor has the inference collapsed everywhere at once. An audit of the largest data-science competition platform across the AI transition found that freshly earned credentials kept most of their predictive value, while stocks of older credentials from assessments that had stopped running lost most of theirs (Yao, 2026). Credentials, on that evidence, are only as informative as the institutions that keep renewing them, which is our point: their value was never stored in the artefact but in an ongoing practice of assessment that someone has to maintain.

Evidential weight has to go somewhere. If artefacts can be manufactured, whatever is harder to manufacture appreciates: participation sustained and witnessed over time, positions defended and revised in dialogue, errors owned in front of people who can tell the difference. Communities observe this. Credentials compress it away.

The cost of leaving this alone is already visible. Admissions committees and hiring panels continue to ask for more artefacts, earlier. More preprints from teenagers (Golden and Purohit, 2023), more repositories, more performance. Each increment selects harder for access and coaching and more weakly for understanding. If the race between manufactured signals and signal-hungry institutions continues, credentials will tell evaluators progressively less about competence.

A problem of formation

When trust in science erodes, the reflex response is informational: more transparency, better communication. We read the problem instead as one of formation, of how research communities shape people. Several literatures support that reading.

Biesta (2009) separates three functions of education: qualification, socialisation and subjectification. Credentialing systems measure the first, and the other two wither for want of indicators. Socialisation into a practice, and the formation of people who care about truth, are what scientific communities run on. Longino (1990, 2002) gives the epistemological version: scientific knowledge is social, produced through criticism, public standards and what she later called a tempered equality of intellectual authority. Honesty and openness to criticism are social virtues. They can only be learned in relation to others.

Situated learning theory describes the mechanism. Lave and Wenger (1991) treat learning as a trajectory of legitimate peripheral participation rather than a transfer of information into individual heads. Newcomers become knowers by doing progressively more consequential work inside a community of practice, in view of people who can recognise quality. Polanyi (1966) supplies the reason codified knowledge cannot stand in for this: we can know more than we can tell. A learner can acquire the definition of confounding without acquiring any sense of when it matters, or when a technically correct answer should still produce doubt. That judgement passes through apprenticeship, by watching another person work with their mistakes and revisions visible. It is also what survives generative AI, because it never lived in artefacts.

We are not proposing that communities replace credentials. Credentials were always an IOU written against community formation, and AI has forced the question of whether the underlying asset exists. Where competence has been formed and witnessed in relationship, a community can testify to it directly. Where it has not, the credential certifies much less than it appears to.

What research communities can do

Moving trust from outputs to formation requires structures in which formation can be seen. Our community is piloting three. They are design principles to test, and we have no evidence yet that they work.

Select for depth rather than legibility

Credentials survive in selection because they are auditable and character is not. Live, unscripted conversation recovers some of that observability. We have moved towards open discussions with prospective members built around questions that have no correct answers. When did something last unsettle you enough to change how you saw the world? What would you do on finding a significant error in a paper you had co-authored? A language model can rehearse an answer to these. What it cannot supply is the thinking behind one: the hesitation and the revision that separate a position held from a position performed. The questions also do formative work whatever the outcome, since they show a candidate the culture they would be entering. Assessment of this kind is subjective, and we treat it as one weak signal among several. A single conversation would reproduce the problem we describe below, since it rewards those already fluent in narrating themselves. So it is not where the decision is made. Entry is provisional, and what counts is what a person shows over the months that follow: whether they

ask questions, own errors and change their mind in front of others. The conversation begins that record. It does not stand in for it.

Build small relational infrastructure and keep it unmeasured

Peer-to-peer learning yields cognitive gains for tutor and tutee alike (Topping, 1996), and disagreement surfaces more readily where psychological safety exists (Edmondson, 1999). Between large project meetings we assign groups of two or three members with no faculty present, whose task is to continue the conversation rather than to produce anything. What did I understand this week that I did not understand before? When did I change my mind? Low-friction messaging channels give these exchanges continuity and recover something of the corridor conversation that formal structures squeeze out. Versions of this already exist in low-resource settings, where formal mentorship is thin and necessity built them. A case-discussion group running on a cheap messaging platform creates the same safety to say “I do not know”. Global research communities should connect to those spaces rather than duplicate them.

Refuse to metricise the alternative

Measurement pulls hard. Left alone it will convert participation into attendance counts and response times, and rebuild the credential-farming machine inside the reform (Campbell, 1979). The distinction that matters is between measurement and judgement. A measure is quantified, comparable across people and therefore optimisable. A judgement is qualitative, tied to its context and made by someone who can be asked to defend it. Formation can be judged, by asking over time whether members change their minds and whether intellectual relationships outlast the tasks that created them, but the moment those judgements are scored and ranked they become targets. Judgement has to stay subordinate to the thing it tracks, or measurement will replace both.

The formation gap

The stakes are unevenly distributed. In many education systems, academic selection remains strictly grade-based. Holistic assessment, exported as the meritocratic alternative, carries a hidden curriculum: it rewards people trained to present themselves as curious and community-minded, and that training is unequally distributed. A researcher formed entirely within a grade-based system may have deep curiosity and real epistemic humility and still find the question “when did something last unsettle you?” unanswerable, because nobody taught them that such experiences count as evidence of anything. This is close to what Fricker (2007) calls hermeneutical injustice: a significant area of a person’s social experience obscured from collective understanding because structural prejudice has kept their group from shaping the shared interpretive resources. The parallel holds because the vocabulary of holistic assessment was built by and for the institutions that export it, not by those formed within grade-based systems, so the interpretive gap falls on the latter and stops them rendering their own experience legible to institutions with power over them.

Informal study groups and peer networks are where many researchers in low-resource settings form one another. That learning is real and institutionally invisible, the inverse of credential farming, in which the institutionally visible turns out to be epistemically empty. A community serious about recruiting beyond its existing networks has to do more than find these researchers. It has to create conditions where relational and practical intelligence becomes visible on its own terms, without fluency in a self-presentation vocabulary nobody taught them.

Limits and risks

We have not solved credential farming, character is not reliably observable, and our selection practices are not demonstrably fair. Nor do we blame individuals for structural incentives. A student who needs elite placement to secure their family's financial stability is not making a mistake by optimising for credentials, and telling young people to stop caring about prestige without changing what prestige buys is neither kind nor realistic. Selection built on conversation carries its own bias, against candidates from places where self-narration is not a trained skill. Communities with strong shared values carry their own pathologies: insularity, and the ease with which "we select for depth" becomes a gatekeeping mechanism much like the prestige system it replaced. We are building in accountability through ethics oversight of any formal study of our community, documented decision-making and rotating leadership. Four of the authors are secondary-school students, two of whom led the conceptualisation and drafting of this Comment, and the profile sketched in its opening sentence is our own. We cannot certify our own formation from the inside, and readers should discount our self-report accordingly. The argument does not require us to be exceptions to it. It predicts that people in our position will be hard to tell apart from those who learned nothing, which is precisely the discrimination we are asking research communities to make. A peer-reviewed critique of academic performance is itself an academic performance, and we accept that tension, since institutional change is usually made by people working inside the structures they want to outgrow.

Conclusion

A practical agenda follows from the argument. Research communities should pilot selection and mentorship structures that make formation observable, and document them honestly, failures included. Funders and institutions that signed the research-assessment reforms should extend them past outputs to evidence of formation: testimony to witnessed judgement, participation that peers will vouch for. Educators should treat generative AI less as a cheating problem to be policed and more as the end of artefact-based inference, and rebuild assessment around process and revision (Swiecki et al., 2022). Global research networks should find the informal learning communities that already exist in low-resource settings and connect to them as peers.

A community offers something a credential cannot: a conversation that goes somewhere neither person expected, after which both know something they did not know before. No institution can mass-produce that. Institutions can decide whether their incentives make it more or less likely, and for most of a

century they could avoid deciding, because the credential system summarised formation well enough. Generative AI has made that option harder to hold. Machines can write the papers; they cannot become someone worth trusting. Telling the difference is now a central task for scientific institutions.

Declarations

Competing Interests

RG serves as a course lead and paid consultant for the Collaborative Institutional Training Initiative (CITI Program) in the development of an online course on artificial intelligence for healthcare providers. The remaining authors declare no competing interests.

Ethical approval:

Ethical approval was not required as this Comment does not report research involving human participants or their data.

Informed consent:

Not applicable, as this Comment does not report research involving human participants.

Author Contribution

R.A.L., L.A.C., D.H., S.M., R.I.R.S., S.R., D.S., A.Q.X. and B.Z. led conceptualisation and wrote the original draft. G.O.A., N.C., R.G., N.J. and A.N. contributed to writing through review and editing. M.M.B. and L.A.C. supervised the project. C.B. handled project administration, including preparation and submission of the manuscript. All authors reviewed and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgements

This manuscript was edited for language and readability with the assistance of a large language model. The model did not contribute to the conception of the argument or to the generation of substantive content. All model output was reviewed and approved by the authors prior to incorporation.

LAC is funded by the National Institute of Health through DS-I Africa U54 TW012043-01 and Bridge2AI OT2OD032701, the National Science Foundation through ITEST #2148451, a grant of the Boston-Korea Innovative Research Project (RS-2024-00403047) and a grant of the Korea Health Technology R&D Project (RS-2024-00439677) through the Korea Health Industry Development Institute (KHIDI) as funded by the Ministry of Health & Welfare, Republic of Korea.

RG is supported by the Johns Hopkins Institute for Clinical and Translational Research (ICTR) and National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH) grant number T32TR004928. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official view of the Johns Hopkins ICTR, NCATS or NIH.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

References

1. Biesta G (2009) Good education in an age of measurement: on the need to reconnect with the question of purpose in education. Educ Asse Eval Acc 21(1):33–46. https://doi.org/10.1007/s11092-008-9064-9
2. Campbell DT (1979) Assessing the impact of planned social change. Eval Program Plann 2(1):67–90. https://doi.org/10.1016/0149-7189(79)90048-X
3. Collins R (1979) The credential society: an historical sociology of education and stratification. Academic, New York
4. Cotton DRE, Cotton PA, Shipway JR (2024) Chatting and cheating: ensuring academic integrity in the era of ChatGPT. Innov Educ Teach Int 61(2):228–239. https://doi.org/10.1080/14703297.2023.2190148
5. Dawson P, Bearman M, Dollinger M, Boud D (2024) Validity matters more than cheating. Assess Eval High Educ 49(7):1005–1016. https://doi.org/10.1080/02602938.2024.2386662
6. Deci EL, Ryan RM (1985) Intrinsic motivation and self-determination in human behavior. Plenum, New York
7. DORA (2013) San Francisco declaration on research assessment. https://sfdora.org. Accessed 27 Aug 2026
8. Edmondson A (1999) Psychological safety and learning behavior in work teams. Adm Sci Q 44(2):350–383. https://doi.org/10.2307/2666999
9. Fricker M (2007) Epistemic injustice: power and the ethics of knowing. Oxford University Press, Oxford
10. Golden D, Purohit K (2023) The newest college admissions ploy: paying to make your teen a peer-reviewed author. ProPublica, 18 May 2023. https://www.propublica.org/article/college-high-school-research-peer-review-publications. Accessed 3 Sep 2026
11. Goodhart CAE (1975) Problems of monetary management: the U.K. experience. Papers in monetary economics, vol 1. Reserve Bank of Australia, Sydney

12. Grant AM, Berry JW (2011) The necessity of others is the mother of invention: intrinsic and prosocial motivations, perspective taking, and creativity. Acad Manage J 54(1):73–96. https://doi.org/10.5465/amj.2011.59215085
13. Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I (2015) Bibliometrics: the Leiden Manifesto for research metrics. Nature 520(7548):429–431. https://doi.org/10.1038/520429a
14. Inoshita K, Omura M, Yamanaka T, Maeda G, Tsuji K (2026) Does AI homogenize student thinking? A multi-dimensional analysis of structural convergence in AI-augmented essays. https://doi.org/10.48550/arXiv.2603.21228. arXiv:2603.21228
15. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG (2016) MIMIC-III, a freely accessible critical care database. Sci Data 3:160035. https://doi.org/10.1038/sdata.2016.35
16. Kasneci E, Seßler K, Küchemann S, Bannert M, Dementieva D, Fischer F, Gasser U, Groh G, Günnemann S, Hüllermeier E, Krusche S, Kutyniok G, Michaeli T, Nerdel C, Pfeffer J, Poquet O, Sailer M, Schmidt A, Seidel T, Stadler M, Weller J, Kuhn J, Kasneci G (2023) ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ 103:102274. https://doi.org/10.1016/j.lindif.2023.102274
17. Lave J, Wenger E (1991) Situated learning: legitimate peripheral participation. Cambridge University Press, Cambridge
18. Liang W, Yuksekgonul M, Mao Y, Wu E, Zou J (2023) GPT detectors are biased against non-native English writers. Patterns 4(7):100779. https://doi.org/10.1016/j.patter.2023.100779
19. Longino HE (1990) Science as social knowledge: values and objectivity in scientific inquiry. Princeton University Press, Princeton
20. Longino HE (2002) The fate of knowledge. Princeton University Press, Princeton
21. Merton RK (1973) The normative structure of science. In: Storer NW (ed) The sociology of science: theoretical and empirical investigations. University of Chicago Press, Chicago, pp 267–278
22. Muller JZ (2018) The tyranny of metrics. Princeton University Press, Princeton
23. Polanyi M (1966) The tacit dimension. Doubleday, Garden City
24. Spence M (1973) Job market signaling. Q J Econ 87(3):355–374. https://doi.org/10.2307/1882010
25. Strathern M (1997) Improving ratings': audit in the British University system. Eur Rev 5(3):305–321
26. Swiecki Z, Khosravi H, Chen G, Martinez-Maldonado R, Lodge JM, Milligan S, Selwyn N, Gašević D (2022) Assessment in the age of artificial intelligence. Comput Educ Artif Intell 3:100075. https://doi.org/10.1016/j.caeai.2022.100075
27. Tamir E, Nairz-Wirth E (2026) Disrupted distinction, weakened signals: generative AI and the changing meaning of academic performance. Paper presented at the European Conference on Educational Research, Tampere, 20 August 2026
28. Topping KJ (1996) The effectiveness of peer tutoring in further and higher education: a typology and review of the literature. High Educ 32(3):321–345. https://doi.org/10.1007/BF00138870

29. Weber-Wulff D, Anohina-Naumeca A, Bjelobaba S, Foltýnek T, Guerrero-Dib J, Popoola O, Šigut P, Waddington L (2023) Testing of detection tools for AI-generated text. Int J Educ Integr 19:26. https://doi.org/10.1007/s40979-023-00146-z
30. Wu E (2026) From production to verification: generative AI, doctoral formation, and the leadership of digital education. Front Educ 11:1890769. https://doi.org/10.3389/feduc.2026.1890769
31. Yao S (2026) Stranded credentials: how a skill-signaling market absorbed generative AI. https://doi.org/10.48550/arXiv.2608.17111. arXiv:2608.17111

Figures

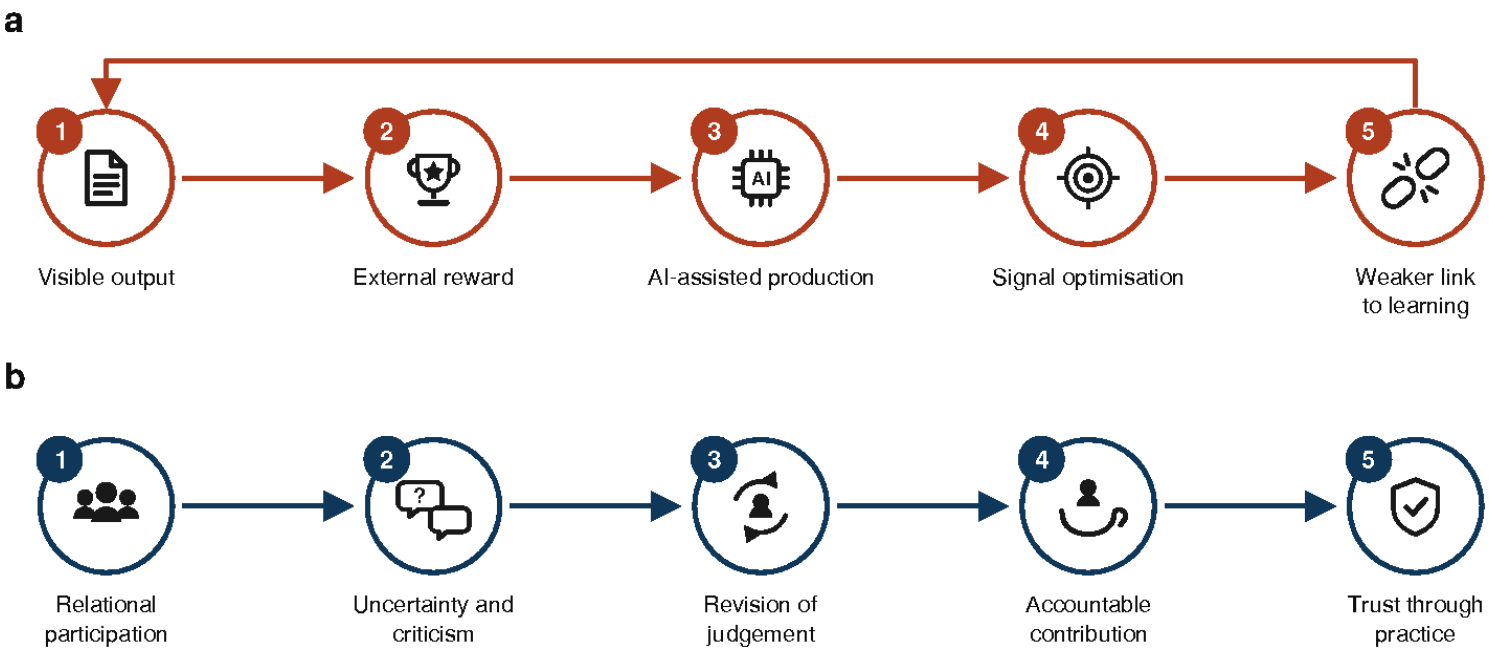


Figure 1

Two ways trust can be built in a research community. **a** The credential-farming cycle, in which visible output earns external reward, AI-assisted production lowers the cost of that output, effort shifts to optimising the signal, and the link between the signal and any underlying learning weakens; the weakened signal feeds back into further output. **b** Formation through practice, in which relational participation makes uncertainty and criticism expressible, judgement is revised in front of others, contribution becomes accountable, and trust accumulates through witnessed practice rather than through artefacts.