

WHO WROTE THIS?

AI AND THE STRUGGLE OVER AUTHORSHIP IN THE ARTS

DICK STROUD

SUMMARY

THE ESSAY ARGUES that debates over AI writing are really arguments about authorship, legitimacy, provenance and trust. It treats the spread of AI in writing as inevitable, not a passing disruption, and suggests that much of the pushback comes from fear that the creative element of writing may not be as uniquely human, scarce or irreplaceable as writers like to believe. AI is tolerated when used for research, transcription, summarising or mechanical support, but becomes morally charged when it shapes judgment, structure, style or creative expression. The essay does not dismiss concerns about hidden provenance, uncompensated training data, economic insecurity or deceptive substitution, but argues that these concerns are often mixed with professional anxiety and aesthetic defensiveness. It also warns that people overclaim their ability to detect AI-written text, making disclosure, process and accountability more important than stylistic suspicion. The conclusion calls for clearer distinctions between acceptable assistance, deceptive substitution and responsible human-AI collaboration.

INTRODUCTION

Kathleen Stock's article, *'The cowardice of the AI plagiarist'* (UnHerd, 3/4/2026) relates the tale of two writers accused of 'cheating' by using AI to write their prose.

One culprit was Alex Preston, the *New York Times* reviewer who admitted using an AI tool to 'expand and smooth' a draft review. The other was Matthew Goodwin, whose book was attacked for errors and supposed machine involvement. Goodwin is an academic who recently fought a by-election for Reform, where he beat the Labour candidate. Not the actions of somebody likely to attract much sympathy from progressives.

Stock's line of argument was uncompromising. Once a writer lets AI write for them, she argued, the computer becomes a ventriloquist and the writer its dummy. It was meant to sting, and it did.

What caught my attention was not her argument, still less the fact that she disapproved – Stock never minces her words. What surprised me was the intensity of her writing, which sounded less like an inquiry than a prosecution. Her language of cowardice, fakery and violated authorship suggested that AI and authorship touch a nerve that can make even a careful and usually balanced writer sound unusually shrill. Something more than a simple argument about standards was driving her narrative.

On the surface, the argument is about what counts as acceptable AI help and when it

becomes cheating or fraud. Beneath that, though, there seemed to be another argument about legitimacy: who is presumed to have done the work, who is granted the benefit of the doubt and who is not. My instinct is that the legacy media have a consistently anti-AI bias, especially about its role in writing and publishing. The reality is more complicated than that, and also more revealing.

AI is tolerated, or at least discussed in a sober and managerial voice, when it behaves like a super search engine: retrieving, transcribing, summarising or helping move information from one place to another. The tone changes when it helps with interpretation, challenging, structuring, reviewing and, above all, creating prose. At that point, the argument ceases to be about productivity and becomes an argument about legitimacy. It's acceptable when limited to support, but expect a stream of insults when it crosses the line into assisting the 'creative magic' of writing.

Even some of AI's defenders concede this boundary. Cindy Yu, writing in *The Times*, argues that writers should not be ashamed to use AI, yet the role she imagines is bounded by assistance with research and the stress-testing of arguments, rather than authorship itself. That is revealing not because her conclusion is the same as that of the sceptics, but because it rests on a similar instinct.

AI is easiest to defend when it stays in the background as part of preparation, checking and development. Emotions are heightened when it appears to be shaping the argument while the work is still presented as the product of a wholly human mind. That anxiety tells us more than the old distinction between tool and author, but it is still not the whole story.

The backlash to AI writing is also driven by worries about concealed provenance, anger at large language models built on unremunerated labour and the assumption that machine-made prose is intrinsically inferior to human writing. Heavens forbid that writers' ire may also be fuelled by fear that their skills and authority are becoming easier to simulate.

WHERE THE ANXIETY BEGINS

The simple question, 'Did the machine write it?' is now too blunt a measure. This boundary is not crossed only when the machine produces the final sentence. Long before that point, unease begins to gather. Most critics appear willing to accept AI's help with the mechanical tasks of managing information. Resistance builds when it encroaches on making judgments and becomes unconditional as soon as it starts creating text.

This helps explain why some uses of AI barely register as morally interesting while others attract language that sounds almost theological in its intensity. Nobody – well, the great majority of people – imagines that locating a source, transcribing an interview or obtaining a rough summary is the heart of authorship. These look like ancillary functions, useful perhaps, but not constitutive of the self that stands behind the writing. Even the phrase 'research tool' performs a kind of moral softening. It places AI in the realm of assistance rather than agency. Acceptable AI is not merely one that helps; its assistance leaves the writer's judgment intact.

The trouble is that this symbolic line is much less absolute than people pretend. Search itself is never neutral, because what one finds shapes what one thinks. Summaries alter emphasis. Reframing a question is already a kind of interpretation. Suggesting counterarguments is already a form of challenge. Proposing an outline is an intervention in structure.

WHO WROTE THIS?

Once this is admitted, the terrain becomes less simple than the reassuring contrast between harmless tool and corrupting substitute. AI can participate in thinking well before it produces any prose.

Linguistic translation is a particularly good test case because it shows how dependent these judgements are on context. Machine translation has been used for years, partly because it predates the current moral drama, partly because most people do not regard it as an intrusion into their creative territory, and mainly because many do not realise it as a form of AI. What matters is not the underlying technology so much as whether people feel that authorship itself has been invaded. A translator does not merely map between two languages; they decide on cadence, connotation, register, formality, idiom and emphasis. In other words, the translator judges. The same technology is suddenly felt differently because the surrounding understanding of the task has changed.

AI used to review books sharpens this argument even further, because a review is meant to show a critic's judgment in their own words. The Preston case is useful here because his offence is difficult to define. A side-by-side reading of his *New York Times* review and Christobel Kent's earlier *Guardian* review does not suggest blanket copying. Naturally, some overlap comes from the novel itself and from plot points that any reviewer would likely mention. The stronger evidence lies in several identical short phrases and, more revealingly, in a strikingly similar structure in one part of the writing. 'Partial appropriation of language and framing' seems a better description than either total plagiarism or pure coincidence.

What made this case significant, then, is not just that a reviewer had used a shortcut. It was that the shortcut appeared in a form of writing where the exact words are part of the judgment. In a review, it is not only the conclusion that matters, but the way the conclusion is reached and expressed. That is why even limited borrowing or AI smoothing could make the judgment feel less authentic. Now, let's back up a minute. It took more than 80 days for the *New York Times* to admit Preston's offence. This does not suggest there was a horde of readers affronted by his writing. The cynic might say the readers benefited from getting the combined review of two reviewers.

The public commentary around Preston was high on indignation and low on analysis. One compromised review became proof of plagiarism, AI hallucinations, copyright violation and the flattening of prose. The more revealing questions were procedural: where exactly did the overlap occur, how strong was it, what did the tool see or retrieve and which editorial safeguards failed?

An AI agent can be conditioned by examples in the immediate prompt, retrieve outside passages behind the scenes and feed them back into a draft and, in some circumstances, use memorised training text. Those are very different things, both technically and ethically. In a case like Preston's, contaminated context or retrieval seems likelier than spontaneous regurgitation, though without knowing his exact prompt and toolchain, that remains an inference rather than a proof. The important point is simpler. Once you stop treating it as a spectacle and start treating it as a process issue, the moral certainty that he did something wrong looks less clear. Alas, such complexities spoil a good story, especially in *the New York Times*.

WHAT AI THREATENS IN THE ARTS

The next question is why this matters so much in the arts. Why should otherwise ordinary forms of assistance suddenly acquire moral weight the moment they appear to drift from retrieval into judgment? One answer is that writers are frightened and defensive, and no doubt there is some truth in that. Another is that critics of AI are simply right because AI-generated prose is self-evidently inferior, and machine use is self-evidently fraudulent. Neither answer will do. The backlash is stronger and more interesting than either caricature allows.

The contentious issue is provenance. In the arts, and especially in writing, the origin of the work is part of what is being consumed. A novel is not merely a sequence of competent sentences; it is a novel by someone. A review is not merely a bundle of assessments; it is an instance of a particular mind at work. A poem, translation, script, lecture, column or essay arrives under a name, and that name does not sit on the cover like a detachable label. It becomes part of the value of the object itself. Rebecca Watson catches the feeling well when she describes the pleasure of reading as a way of animating somebody else's carefully wrought imagination. Knowing that there is somebody at the far end of the prose is not an optional extra. It is part of what the reader thinks they have bought. Well, at least readers like Rebecca Watson.

This is why so much public language around AI in the arts turns so quickly to words like fake, ventriloquism, disguise, concealment and betrayal. The concern is not only that a machine may have helped, but that the visible relation between the work and the person claiming it has become tarnished. Once that uncertainty appears, its cultural value begins to wobble. One may still admire the output. One may still find it useful, moving or clever. But the basis on which admiration was first offered has shifted, and people feel that shift very strongly. The inability to discern human from AI-generated text does not make the distinction irrelevant; it sharpens the demand for disclosure.

The wider arts examples make the point more clearly because they show that the problem is not peculiar to books. In music, a 2025 Deezer-Ipsos survey reported that most listeners could not reliably distinguish AI-generated songs from human-made ones, yet a strong majority still wanted such tracks labelled as AI-generated. The Writers Guild of America drew its own line with striking bluntness: AI is not a writer, and AI-generated material does not count as literary material under its agreement. In both cases, the issue is not only whether the output passes. It is whether audiences, institutions and collaborators are entitled to know what sort of thing they are dealing with, and how it came into being.

A second anxiety is both economic and moral. Critics object that large-language-model prose is built on the unremunerated labour of writers whose work helped train the models, even though they are neither credited nor compensated. This is not quite the same as saying that any given AI output is plagiarised, which is a narrower and often more legally loaded claim. It is something broader and, in its own way, more unsettling. Watson's account of reading the suspect AI-shaped novel *Shy Girl* makes the point in miniature. Her resistance to the book is both aesthetic and personal. She notes that her own novels are available on LibGen, a large pirate library of books and articles that has reportedly been used as training data for some AI models. The machine's fluency is therefore felt not as an innocent novelty, but as something that relies on unpaid human labour.

WHO WROTE THIS?

The Christie's AI art controversy made the same complaint on an industrial scale. More than 3,000 artists objected to the auction house's sale of AI-related works on the grounds that many of the models had been trained on copyrighted material without permission, even as Christie's pressed on and later celebrated the auction's financial success. The grievance is that art has become an industrial process: hidden inside a black box, monetised at scale and separated from the accepted ways in which artists have drawn on each other's work, answered it and gradually changed a tradition over time. At this point, you are either frantically nodding or thinking, 'what a load of artistic nonsense.'

Another anxiety is philosophical rather than legal. Large language model-generated text is often dismissed as nothing more than the fusion of other people's ideas. It does not think, critics say; it recombines. There is force in that objection, but it also opens a trapdoor beneath the critic's feet, because the awkward truth is that all writing is, in some degree, recombinant. No serious writer begins from absolute originality. Writers evolve from what they have read. Their language is shaped by others' language, their arguments are rarely original, their sentences are shaped by remembered cadences, their preferences by admiration and aversion, their sense of the sayable by all the voices they have heard or read. The literary self is formed through inheritance, imitation, struggle, correction, theft, resistance and selection.

The more serious question, then, is whether AI borrowing differs from human influence in kind, or chiefly in scale, speed, opacity and the absence of accountability. Human writers also draw on ideas and language they have encountered before, but they do so as people making choices. They decide what matters, what to stress and what to leave out, and they can be questioned about those choices. What they write is tied to their own life and reputation, something that echoes back to the reader, who is buying both the prose and the person.

Machine recombination may produce superficially similar effects, but it does so without owning its inheritances in any recognisable human sense. To say that all writing is recombinant is not to say that all recombination is culturally equivalent.

The argument does not end with the issue of provenance and extraction, far from it; there is also the more sweeping condemnation that machine-generated prose is, by nature, thinner and less valuable than ordinary human writing. Watson does not simply sneer at the prose of *Shy Girl*, she describes it as full of recurring three-part lists and mechanical over-insistence, then pauses long enough to ask whether she is identifying the influence of AI or merely indulging her own snobbery. That hesitation matters. It shows that someone may genuinely sense something mechanical in a piece of writing without that sensation being reliable evidence by itself. Let's remember they are fallible humans, not machines.

A lot of AI writing today sounds too polished, explains too much, avoids real feeling or risk, sounds more confident than it should and tries to sound important rather than actually being meaningful. Hold on, that last sentence could have been extracted from zillions of book reviewers' musings. But there is an important difference between saying that much AI prose is bad and assuming that machine-generated prose must be bad in principle.

Watson is arguing about more than whether AI borrows too much or writes badly. She is also saying that part of a novel's value lies in the fact that a person struggled to write it. The finished book matters partly because of the human effort, judgment and oddness behind it. AI fiction, in her view, appeals to a culture that wants the result without caring much about the long process that makes the result worth having. That seems to me like a huge leap of value judgment, but let's move.

It may sound romantic, but it points to something real. Writing is not just the act of putting words on the page. It also includes thinking things through, starting badly, going back over what you wrote, throwing parts away and slowly building up judgment over time. A tool advertised as a way of skipping that process is therefore experienced not just as efficient but as anti-artistic. Maybe this argument can be summarised as ‘no pain, no gain’?

Grayson Perry is useful as a counterpoint because he complicates the emotional map. He is hardly a naive evangelist for technology, yet he has said that he is not especially troubled by AI using his work, partly because the real value lies in the original artwork itself, and partly because art has always involved taking and reworking what came before.

He was also candid enough to note that he speaks from a relatively secure position. It suggests that reactions to AI are shaped not only by principle but by what, exactly, the artist thinks is at risk. A painter whose market depends on unique physical objects may feel less exposed than a writer whose entire medium consists of reproducible language.

Put all this together and a more complex picture begins to emerge. The backlash to AI in writing and the arts is driven by linked anxieties: that AI obscures provenance, that it draws on uncompensated labour, that it recombines without accountable judgment and that the prose or imagery it produces is intrinsically inferior. None of these anxieties is wholly irrational. Together, they create a climate in which AI becomes more than a tool or even a threat. It becomes a symbol of a world in which the visible relation between effort, judgment, originality and reward is dissolving.

MAXIMUM CERTAINTY, MINIMUM EVIDENCE

Unfortunately, judgment about this subject often outstrips its own powers of proof. People sound very sure about what AI prose is, what it feels like and what it has done. Yet the moment one asks how reliably these claims can be established from reading the finished text alone, certainty begins to look less secure than the rhetoric suggests. That is not a side issue. It is central to why the subject has become so morally charged. Much of the heat comes from the fact that suspicion is easy to feel and hard to prove.

At this point the argument takes an awkward turn because two quite different claims are often run together. One is the weaker claim that some AI-generated prose often carries habits that can make it feel recognisable. The other is the much stronger claim that one can reliably attribute a particular text, or a particular degree of machine involvement, from the text alone. This second claim is much less secure than public rhetoric likes to imply.

The evidence is not altogether one-sided. In simplified settings, human beings can sometimes do a little better than chance at distinguishing fully AI-generated text from fully human text. But those results flatter the practical reality because the task is usually easier than it is in ordinary life. In a blind field test at the University of Reading, researchers inserted wholly AI-written submissions into a live university examination system and found that 94% went undetected; those submissions also received marks that, on average, outperformed comparable student work. A 2025 study in *Advances in Simulation*, designed around five more realistic authorship conditions ranging from fully human to fully AI-generated, found that human raters achieved only 19% accuracy. That is not a picture of a public confidently spotting fraud. It is a picture of uncertainty disguised as certainty.

The weakness is not confined to crude student prose. In a 2024 *Scientific Reports* study,

WHO WROTE THIS?

non-expert readers performed below chance when asked to distinguish AI-generated from human-created poems. More strikingly, the AI poems were judged more likely to be human-authored than the actual human poems and were rated more favourably on qualities such as rhythm and beauty. Whatever one thinks this proves about poetry, it does not support the comforting idea that machine text is always obviously flatter, emptier or easier to detect than ordinary human writing. It suggests something more awkward: readers may be relying on shared but mistaken heuristics when they decide what sounds human. At this point, the image of a French champagne judging competition came to mind, with looks of astonishment when the judges discovered the winner came from Australia, not the Côte des Blancs.

Detector software is only slightly more accurate, and only under carefully managed conditions. A 2025 study in *Acta Neurochirurgica* reported strong results when detectors were asked to distinguish AI-generated from human-written neurosurgery texts. But the same study also emphasised that none of the detectors achieved full reliability and that false positives remained a significant risk. That caveat matters more than the headline number, because real disputes do not concern two idealised piles of text. They concern this particular essay, by this particular person, in this particular state of drafting and revision. Multiple authorship makes the attribution of text much harder.

These difficulties explain why official guidance in higher education is moving away from fantasies of detector-led certainty. Newcastle University states plainly that AI-generated text cannot be reliably detected and does not recommend automatic AI checkers. The Joint Information Systems Committee's guidance makes the same point in more cautious language: detection tools may have a part to play, but they are not a full solution, they are easy to evade, and they cannot conclusively prove that text was written by AI. The common drift is unmistakable. Where institutions have thought seriously about the matter, they have not emerged with greater confidence in spotting AI.

What one sees in Kathleen Stock can also be seen, in a more theatrical register, in Andrew Orłowski's recent Telegraph column on science and AI. The piece is not without evidence, pointing to studies suggesting that heavily automated papers are often weaker, that the flood of polished output makes it harder to separate signal from noise, that large language model editing can soften or redirect intended conclusions and that AI-shaped reviewing may reward conformity over insight. But those findings are quickly made to indicate a much more serious problem.

From there, the column moves with remarkable speed to claims about cognitive decline, institutional degradation and civilisational decay as though the existence of some genuine problems had already vindicated the most sweeping diagnosis available. This is the pattern that commentary on AI so often produces: not always an absence of facts, but a collapse of proportion between fact and verdict. The register becomes prosecutorial before the case has been fully made, as in the case of Stock's article. What drops out are the very things the subject most requires: scale, mechanism, degree, counterexample and a clear distinction between what has been demonstrated and what is merely inferred. I guess you could say these weaknesses demonstrate that the critics are human.

That does not mean the sceptics are wrong about everything. There are plainly stylistic habits that current systems often produce: explanatory smoothness, a tidy insistence on balance, a fondness for symmetrical framing, an eagerness to resolve tension rather than sustain it. But the existence of such tendencies is not the same as a reliable method of attribu-

tion. Human beings can write like that, too, and often do. And once a writer edits, reorders, cuts, roughens or intensifies AI-produced material, the leap from ‘this has a certain feel’ to ‘I know what happened here’ becomes even less defensible. I have probably changed about 10% of this AI-generated text – have you spotted my interventions?

This is why I keep returning to the same concern: that discussion of this subject often combines maximum aesthetic certainty with scant evidence. What people often have is not a dependable forensic method but a set of impressions and stylistic stereotypes. Some of those impressions are sensible. Some are absurdly overconfident. Many are socially and politically inflected. A text written by a student, a politician, a pundit or a critic already suspected of shortcut-taking may attract suspicion more readily than the same prose under a different name. Where proof is hard to secure, character judgments rush in to fill the gap.

That is one reason why disclosure and provenance are becoming more, not less, important. If the finished text cannot reliably settle the question of origin, then arguments about authorship are pushed away from the surface of prose and back towards process: drafts, notes, acknowledgements, declared method, editorial norms and institutional trust.

THE INSTITUTIONS AT RISK

Different institutions are not defending the same thing, so they do not react in the same way. A newspaper column is not a novel, and neither is it a screenplay, a translated poem, a piece of journalism or an audiobook performed in the author’s own voice. The confusion begins when all these activities are discussed as though they were interchangeable examples of writing, and all objections as though they sprang from one emotional source. They do not. Each institution feels the pressure of generative AI at the point where its own claim to legitimacy is most vulnerable.

The divergence between the workplace and the university is the clearest example. Employers increasingly value fluency with AI tools not because they have undertaken a philosophical revaluation of authorship, but because businesses buy outcomes. Universities face a different problem. A degree is not merely a record of what was submitted but a claim about whose mind acquired what capacity. The institution must be able to draw some defensible line between help and substitution, or the credential starts to devalue. The workplace demands visible competence with tools. The university still has to certify a more personal kind of achievement.

That tension will deepen as AI becomes embedded in ordinary software, making it harder to draw a clean boundary between using AI and not using it. Academics fret that hidden assistance threatens the validity of their credentials, while competence with AI tools is mandatory for consultancy companies. Much of what appears to be hypocrisy is better understood as a structural divergence between institutions that protect different goods.

The publishing industry might tolerate the use of AI to drive down costs while erecting safeguards to keep it away from anything deemed creative. Low-level tasks that once would have been offshored to India, such as translation, metadata generation, audio production, proofreading and other forms of process efficiency, are now handled by AI. At the same time, it continues to sell books under the signs of named authorship, singular voice, editorial curation and the aura of originality.

I want to return to the case of Goodwin because his opinions are anathema to much of the

WHO WROTE THIS?

media class, yet he is not remotely fringe in the country at large as was shown by his party's spectacular results in the recent local elections (May 2026). In that setting, AI can become a way of saying not just that a text is defective, but that the person behind it is somehow fake. Where proof is uncertain and prior dislike is strong, accusations about machine use can become instruments in a wider contest over legitimacy.

The same point helps explain why Stock's harsher tone mattered to me in the first place. She is not a lazy culture-war scold. She is perfectly capable of careful distinctions, and that is why the change in register was revealing. What I heard in her article was not merely hostility to a tool, but alarm at a wider unravelling. If writers can use machines to shape the language in which judgment appears, and if readers cannot reliably tell when this has happened, then something central to literary and intellectual authority begins to feel unstable.

That is why the argument cannot be dismissed as a simple panic, though panic is sometimes present. It touches livelihoods, certainly, but it also touches hierarchy. For a writer, cultural authority depends on the relative scarcity of polished prose expression and on the assumption that the ability to produce it reflects a distinctive mental discipline. To be sure, the material rewards of being a writer accrue only to the top 1 per cent, with most labouring below minimum-wage rates. If, let me say when, generative AI creates the same quality of prose cheaply and at scale, then the reward of status attached to those creative abilities vanishes. This does not make serious writing worthless. It does mean that AI has exposed how much of our cultural sorting depends not only on originality in the highest sense, but on the social power of sounding intelligent in public.

WHAT A WRITER MEANS NOW

The question is no longer just whether AI can write acceptable prose, it already can. Nor is the question simply whether some uses are deceptive, though some plainly are. The deeper issue is what happens when output becomes easier to simulate than originate, and when the relation between style and authorship can no longer be inferred with the old confidence. In such a world, cultures of writing will have to rely more heavily on declared method, provenance, editorial norms and explicit rules about what sorts of assistance are permitted where. They will also have to give up a fantasy of purity. Most real cases now sit somewhere between untouched human expression and fully machine-generated output. The challenge is to decide which forms of collaboration preserve responsibility and which dissolve it.

The sensible response, then, is neither blanket denunciation nor airy indifference. It is to become more discriminating. Search is not the same as structure. Translation is not the same as literary voice. Tidying is not the same as reviewing. Brainstorming is not the same as substitution. Hidden co-authorship is not the same as declared assistance. Writing about AI is least useful when it treats every dubious case as proof of the technology's inherent badness and most useful when it asks procedural questions instead: what the system was shown, what it retrieved, where the overlap actually lies and which safeguards failed. Once those distinctions are made, the present argument becomes less hysterical and more exact, though no less important for that.

I began with Kathleen Stock because her unusually heated article was the thing that made me feel this problem rather than merely observe it. I end in a slightly different place. What I first took to be a simple anti-AI feeling turned out to be a tangled mix of principled objec-

DICK STROUD

tions, economic grievances, aesthetic prejudices, class instincts and institutional self-defence. Some of those objections are stronger than others. Some are plainly overdrawn. Some rely on assumptions that need testing rather than repeating. But the dispute itself is not frivolous. It is a struggle over what counts as writing, who gets to claim it and what readers, listeners and audiences are entitled to assume when a human name appears above a text. The quarrel is not, in the end, about whether machines can write. It is about what we mean by a writer once they do. And let's not forget the final arbiter in the argument – what the customer is willing to purchase.