AI Could Have Written This:
Birth of a Classist Slur in Knowledge Work

Advait Sarkar
advait@microsoft.com
Microsoft Research, University of Cambridge, University College London
United Kingdom

Target of judgement

AI use confirmed?	Person is judged	Artefact is judged
AI used	**Shaming**: person is judged negatively for using AI	**Hard blaming**: artefact is judged negatively for using AI
AI use unconfirmed	**Allegation**: person is judged negatively for potentially using AI	**Soft blaming**: artefact is judged negatively for potentially using AI
AI not used	**Slur**: person is judged negatively through comparison to AI	**Disparagement**: artefact is judged negatively through comparison to AI

Figure 1: An incomplete taxonomy of AI shaming terms. More detailed definitions are given in Section 2.

Abstract

AI shaming is a social phenomenon in which negative judgements are associated with the use of Artificial Intelligence (AI). This includes comparing someone's work with AI-generated work as a means of disparagement, voicing suspicion or alleging that someone has used AI to undermine their reputation, or blaming the poor quality of an artefact on AI use. Common justifications of AI shaming include recourse to AI's societal harms, its technical limitations, and lack of creativity. I argue that, more fundamentally than any of these, AI shaming arises from a class anxiety induced in middle class knowledge workers, and is a form of boundary work to maintain class solidarity and limit mobility into knowledge work. I discuss the role of AI shaming in protecting the privileged class of knowledge work and its attendant harms.

CCS Concepts

• **Human-centered computing** → **HCI theory, concepts and models**; Interaction design theory, concepts and paradigms; **Collaborative and social computing theory, concepts and paradigms**; • **Computing methodologies** → **Philosophical/theoretical foundations of artificial intelligence**; *Discourse, dialogue and pragmatics.*

Keywords

class identity, sumptuary laws, protectionism, epistemic injustice, colonialism, sociotechnical imaginaries, moral panic

1 Introduction

This paper interrogates a bundle of social phenomena that associate negative judgements with the use of Artificial Intelligence (hence **AI**). This bundle includes comparing someone's work with AI-generated work (regardless of whether it really involved AI) as a means of disparagement, voicing a suspicion that someone has used AI as a means of undermining their reputation or integrity, blaming the apparently poor quality of an artefact on AI use, or simply shaming someone for using AI (Section 2).

Such slurring, disparagement, allegation, blaming, and shaming is typically supported by one or more of a set of common arguments: AI output is of poor quality, AI cannot be creative, AI causes societal harms, and so on (Section 3).

In this paper, I argue that these phenomena in fact arise from a class anxiety induced in knowledge workers, and are a form of boundary work performed by knowledge workers to maintain class boundaries and limit mobility into knowledge work (Sections 4 and 4.1). In this reading, the increasingly common refrain "AI could have written this" is not so much a pithy taunt, but rather a classist slur, indicative of wounded and anxious privilege. Moreover, it is complicit in the systematic exclusion of underprivileged groups from entering the class of knowledge professionals.

Not only are boundaries to knowledge work constructed through the aforementioned discursive blaming and shaming practices, but they are also crystallised in the form of organisational and institutional rules and codes of ethics which delineate permitted uses of AI from those forbidden (Section 4.2). Underlying these codes is

knowledge materialism: a latent notion of ownership over knowledge work, which is not confronted directly but instead modulated into a morality of knowledge work wherein actions that threaten the material prosperity of knowledge workers are deemed immoral.

Unlike the pressure put on the working classes by industrialisation, which found its ultimate salvation in the propositions of Marx, Generative AI puts pressure on a privileged class by threatening to erase the moat it has so carefully dug between itself and those less privileged. Thus, to understand AI shaming as a response to the effects of Generative AI on the class identity of knowledge workers, it is less instructive to invoke the industrial revolution (though this analogy is still useful to understand changes in the nature of production). Instead, I briefly suggest alternative historical analogies to AI shaming and institutional codes such as sumptuary laws and protectionist policies (Section 4.3).

The proliferation of formal and informal shaming practices induces a cavalcade of societal harms, including psychological disorders, racial discrimination, and chilling effects (Section 5).

The implications of AI shaming are complex, raising the questions of how it might be resisted, whether it serves useful functions that ought to be preserved, and whether the erasure of shaming will aid in fulfilling the potential of Generative AI to improve social mobility or merely reconfigure class boundaries (Section 6). However, one clear implication for our research community – an exceedingly privileged class with steep boundaries and deep incentives for maintaining them – is to become reflexively conscious of AI shaming, to avoid and resist it when possible, and to consider whether our discourse and institutional codes are merely accessories to an exclusive and inequitable enaction of class solidarity.

2 A Provisional Taxonomy of AI Shaming Phenomena

Negative associations with AI can manifest in a number of subtly distinct and overlapping phenomena. In particular we are interested in cases where an individual or group of people are associated with one or more artefacts that may or may not have been generated in a workflow involving AI. These phenomena vary along the following principal dimensions: (1) whether AI is known to have been used, and (2) whether the person or the artefact is the target of the negative association. To distinguish these different cases, I propose the following provisional taxonomy:

(1) **Shaming**: drawing attention to someone's confirmed use of AI with a negative moral judgement or reflection on that person.
(2) **Allegation**: raising suspicion of someone's unconfirmed use of AI with a negative moral judgement or reflection on that person.
(3) **Slur**: comparing someone's work which is confirmed *not* to be using AI with AI-generated work, as a means of critique or negative reflection on that person's abilities.
(4) **Hard blaming**: critiquing the quality of an artefact that involved a confirmed use of AI, attributing its perceived low quality directly to the use of AI.
(5) **Soft blaming**: critiquing the quality of an artefact with no confirmation of the use or non-use of AI, hypothesizing a

link between its perceived low quality and the potential use of AI.
(6) **Disparagement**: comparing the quality of a confirmed non-AI artefact with AI-generated artefacts, as a means of critique or negative reflection on that artefact.

These definitions are summarised in Figure 1. For the remainder of this paper, "shaming" will be used as an umbrella term that encompasses all these phenomena. As we will see, real-world episodes of shaming often involve multiple phenomena. For example, it is common for both the person and the artefact to be judged negatively in the same discourse, so depending on the status of known AI use, we encounter both shaming and hard blaming, or allegation and soft blaming, or a slur and disparagement simultaneously.

This taxonomy is necessarily provisional, incomplete, and simplified for the purposes of this paper. It could be further divided and extended; for instance it may be useful, where the use of AI is unconfirmed, to distinguish between cases where discussants are genuinely interested in discovering whether AI use is involved (so that the discourse may potentially be "hardened" into hard blaming and shaming), and cases where the actual use of AI is not important and the underlying intent is to slur or disparage.

3 A Provisional Taxonomy of Arguments in Support of AI Shaming

Behind the complex collection of AI shaming phenomena is an even more complex constellation of ideas and arguments that justify a negative association with AI. The arguments fall loosely into one of three primary categories: arguments from society and materiality, arguments from technical basis, and arguments from creativity. These are invoked to support secondary categories such as "AI cannot be creative".

An incomplete taxonomy of arguments invoked to support AI shaming is given in Figure 2. The arrows in this figure may be read as "implies" relations. For example, "AI makes errors" **implies** "AI output is of poor quality". In common discourse, primary and secondary categories are often intermixed, and often a leap is made from a primary category directly to the conclusion that a negative judgement is warranted (performing the "transitive closure" of the arrows in the diagram; thus arrows could be drawn directly between primary arguments and the final judgement). Implications are also made beyond those directly represented, e.g., "AI cannot transcend training data" **implies** "AI cannot be creative", and "Creativity requires struggle / labour" **implies** "AI degrades work", but arrows for these are excluded as are transitive closure relations for clarity of the figure.

As with the taxonomy of AI shaming phenomena, this taxonomy of supporting arguments is necessarily provisional and incomplete, serving mainly to establish a context for the subsequent discussion in this paper, rather than claiming to be definitive. As with the shaming taxonomy, a more comprehensive analysis based on a large and replicable shaming corpus would make for interesting future work. These categories of shaming arguments are now illustrated with examples.
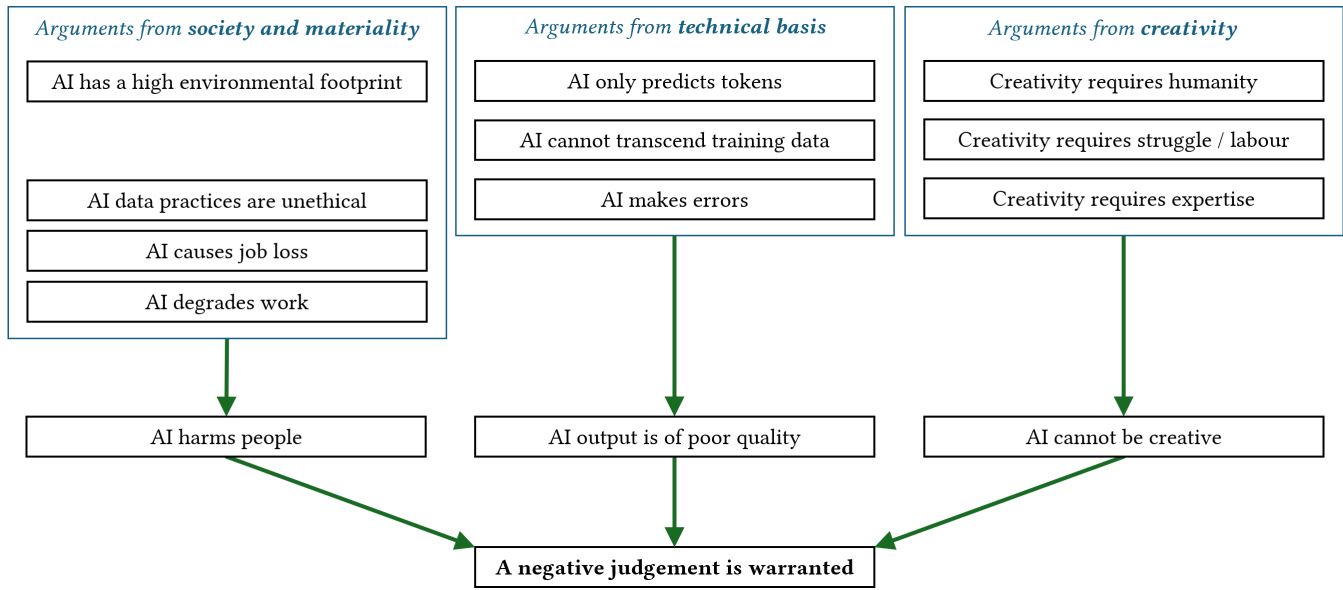
Figure 2: An incomplete taxonomy of arguments in support of AI shaming. Arrows may be read as "implies" relations, e.g., "AI degrades work *implies* AI harms people *implies* A negative judgement is warranted".

Arguments from creativity. Nowhere is the narrative of loss and deterioration more intense than in the application of AI to art. Humanity is posited as an essential element of creativity and art: *"there is art in anything done by a human, and when humans are removed from a task, so is the art"* [38]. The absence or retreat of human involvement in works of art and literature is often cited as *ipso facto* undesirable, and thus justifies shaming. For instance, *"[...] the prospect of a machine capable of writing a book seems almost unbearably sad. An assault on a profound aspect of what it is to be a human being. [...] to believe a machine is capable of producing literature is to misunderstand what literature is [...] A feeling reconstructed by a machine [...] has no human hand to reach out to us [...] even a conscious machine could not really make art"* [44].

Some theorise that art is the result of *"choices at every scale"*, and that there is no *"opportunity to make a vast number of choices using a [e.g.] text-to-image generator"*; as a result those who think AI generated content is acceptable are tasteless novices, uncultivated dilettantes, *"people who think they can express themselves in a medium without actually working in that medium"*, complicit in *"a fundamentally dehumanizing technology"* [16]. A related and extreme form of this argument is that art involves struggle that originates in the divine struggle of creation: *"The day of rest [in the Book of Genesis] is significant because it suggests that the creation required a certain amount of effort on God's part [...] some sort of artistic struggle"*; and because *"ChatGPT rejects any notion of creative struggle [...] by mechanising the imagination it renders our participation in the Act of Creation as valueless and unnecessary [...] that songwriter [...] who is using ChatGPT to write his lyrics [...] is participating in the erosion of the world's soul and the spirit of humanity itself and to put it politely should f**king desist if he wants to continue calling himself a songwriter"* [14].

A case where an artist claimed copyright of AI-generated work (Allen v. Perlmutter, No. 1:24-cv-2665 (D. Colo. 2024)) [11] has been the subject of much shaming. One commentator rejects the artist's characterisation of AI as a tool such as a brush or a camera through the creativity-requires-struggle/labour argument [50]: *"a brush or a camera does not create anything for you. You, behind the camera and the brush, still have to make the art. This [AI] does it for you. [...] You still have to be talented in order to use the tools like a brush or camera to create art. When it comes to AI, you don't have to have talent at all."* Notably, the appeal to labour extends beyond the labour required in the act of creation, to include the labour expended in acquiring the skills of artistry: *"I am not a talented artist at all. I took two years of art courses and only marginally improved. That is something you have to work at for years. With AI art, I type in a single sentence, and it does it for me."* The shaming continues with appeals to work degradation and job loss: *"the problem comes when you try and copyright it, you try and sell it, or you try and deceive people with it, or you try and use AI art tools as a means of pushing out real artists."*

Arguments from society and materiality. Other common accessories to AI shaming are the ideas that automation leads to job loss, or to the deterioration in the quality of experience of a profession. These are, of course, all lay recapitulations of Braverman's work degradation thesis [10], in many ways itself a reinvention of the Marxian fable for the 20th century. For instance: *"as an AI annotator [...] I was very aware of the irony of my situation. Large language models were supposed to automate writers' jobs. The better they became through our work, the quicker our careers would decline. And so there I was, feeding our very own Moloch."* [24]. Commentators recoiled in horror at the suggestion that *"your grandchildren will be the last generation to read and write"* made (apparently without

irony) by an evangelist of AI-generated videos [48]. Besides work degradation, arguments from society and materiality also invoke the disputed use of copyrighted resources and the attendant social impacts, and the environmental footprint of the high material and energy costs involved in AI system development and operation.

Arguments from technical basis. The technical underpinnings of AI, such as reductionist descriptions of its mechanism in terms of next-token prediction, or its alleged delimitation by its training data, are often cited as reasons for the perceived inferiority of AI-generated content. Corollaries include its disembodiment and lack of human-like reasoning and understanding. For instance, an article by Bender [8] begins by restating that *"LLMs are nothing more than statistical models [...] set up to output plausible-sounding sequences of words*, following through by shaming users for using a chat interface for information retrieval: *"The chatbot interface invites you to just sit back and take the appealing-looking AI slop as if it were "information". Don't be that guy."*

AI output is of poor quality. The perceived inferiority of AI-generated content is also mobilised directly in support of shaming. For example, consider the derogatory epithet *"slop: a term of art, akin to spam, for low-rent, scammy garbage generated by artificial intelligence"*, which threatens to *"crowd out human creativity and intentionality with weird AI crap"* [54]. The contemptible culprit-winners of the "slop economy", then, are those who constitute *"a thriving, global gray-market economy of spammers and entrepreneurs, searching out and selling get-rich-quick schemes and arbitrage opportunities, supercharged by generative AI"* [54]. Not surprisingly, as we will revisit, the ire is drawn at entrepreneurs from the Global South, whose ungrammatical prompts are often paraded as exhibits in slop vilification (e.g., [54]). Commentators complain of a "mental slop tax" – the burden of diligence in detecting (and then presumably ignoring) AI-generated content: *"Those of us who have no interest in engaging with slop find ourselves performing a new kind of labor every time we go online [...] We look twice to see whether a "farmhouse" has architecturally nonsensical windows, or whether an X account posts a suspiciously high volume of bot-ishly generic replies, or whether a Pinterest board features portraits of people with too many fingers."* [15]. Commentators claim that as a result, *"the human internet is dying"*, using incendiary and value-laden metaphors such as *"infecting"*, *"polluting"*, and *"bastions [...] being overrun"* to describe the introduction of AI-generated content to the internet [39].

A typical shaming post on social media [23] combines many of the themes discussed so far; an appeal to the limitations of the algorithm: *"they [AI models] work out what the most likely next word will be. That's it."*, a contrast to humanity: *"it's not a brain, it's not like a human"*, appeal to the metaphysical "retrodiction" thesis, i.e., the idea that novelty cannot arise from reuse of materials:[1] *" it's not capable of being better than the information it's fed. And what's it eating? [...] AI-generated crap"*, its disputed practices regarding intellectual property: *"it's also reading my friends' books, without paying"*, its financial and environmental footprint: *"just one iteration of ChatGPT cost $63m AN HOUR for THREE MONTHS to train"*, raising the spectre of job loss: *"Don't tell me you've made a fun new*

tool "for when you don't have a content designer", because agencies will believe you - and they'll try and cut costs by cutting headcount"*, ending with light shaming: *"nobody is really engaging their brains. They're just accepting these tools at face value"*.

———————————

It is not the purpose of cataloguing these arguments to refute them or deny their validity. They may well be truthful and justified reasons for imparting a negative judgement on an artefact or person by association with AI. However, I will suggest that while possibly true, these arguments cloak a deeper underlying motivation for shaming, of which may the shamer may not even be conscious.

This motivation is that AI is seen as a threat. The particular form of threat that AI poses is to an identity possessed by the shamer, that they covet, that they value materially and spiritually, and that maintains its value through exclusivity. This is the class identity of the professional; most commonly that of the knowledge worker. The work degradation and job loss arguments come close to revealing this motivation, but there is a subtle distinction: those arguments emerge primarily from the experience of *individuals*; from their perspective, jobs are lost to bogeymen of automation and corporate rapacity, or an anonymous, undifferentiated mass of unskilled opportunists. In contrast, the latent motivation for shaming precipitated by the threat to identity emerges from processes by which a *group* forms an interior solidarity – a class solidarity – to defend itself and consolidate its control over resources from a well-defined outgroup. In particular, AI enables an outgroup of less-enculturated workers (the shamed), those without access to privileged education and professions, albeit those with the affinity for technology and a certain entrepreneurial grit, to competently perform knowledge work that was heretofore the preserve of a privileged elite (the shamers).

These arguments are always raised by those who can "pass" in society without AI assistance, those whose training and enculturation furnishes them with access to, and equips them to perform, exclusive and valuable knowledge work professions. As those elites with "notational privilege" determine the boundaries of creativity in law [58], so those with what might be called *epistocratic privilege* delineate the boundaries of acceptable AI use in knowledge work.

4 Demarcating Knowledge Work: AI Shaming as Boundary Work

AI shaming is very clearly a form of boundary work [25], originally defined in the context of professional scientists as *"Construction of a boundary between science and varieties of non-science"*, which *"is useful for scientists' pursuit of professional goals: acquisition of intellectual authority and career opportunities; denial of these resources to "pseudoscientists"; and protection of the autonomy of scientific research from political interference."* It operates by adopting *"an ideological style found in scientists' attempts to create a public image for science by contrasting it favorably to non-scientific intellectual or technical activities"*. Foucault's theory of disciplinary power [21] provides the diffuse backdrop against which boundary work becomes worthwhile; institutionalised demarcations (of madness from sanity, of criminality from innocence, of science from non-science, etc.) depend on the coordinated enaction and performance of boundary work which enable arrangements of power to be brokered.

---

[1]The idea that reuse precludes novelty is not consistent with the way many artistic domains conceive of reuse [58].

AI shaming demarcates good from bad knowledge work, acceptable from unacceptable knowledge work, ethical from unethical knowledge work, etc. At its most extreme, AI shaming sets the boundaries of knowledge work itself, defining what does or doesn't count as knowledge work, denying the very ontological status as AI-generated artefacts as work (e.g., the aforementioned commentary on Allen v. Perlmutter). There are many social and material incentives for someone to perform this type of boundary work. For instance, it might raise their own self-esteem as a knowledge worker with epistocratic privilege; showcase their professional status as a knowledge worker with epistocratic privilege; and justify and protect their income – under perceived threat from AI – by supporting the cultural conditioning of AI as immoral and undesirable.

The academic library JSTOR has developed *"JSTOR Daily Sleuth—a game that tests your ability to detect whether the title of an academic research paper is authentic or generated by AI"*, on the basis that *"AI-generated research has presented academic libraries with a new information literacy agenda"* [70]. Note the implied dichotomy between "authentic" and "generated by AI". This can be viewed as the analgesic response of a powerful yet threatened institution with reasons to protect its boundaries as an authoritative vault of knowledge. The aim of the game, quite transparently, is to build the muscle of boundary work. It is not, for example, a game that trains you to detect faulty claims or false information, but merely to detect when paper titles appear to be AI-generated. The game's morally saturated title constructs the narrative of a baddie to be "sleuthed" out, a crime to be solved. It is telling that, in the article, the game is depicted as part of a university installation with the pugilistic provocation: *"Can You Beat the Bots?"*

Boundary work is a war waged primarily by those belligerents who "possess" knowledge work against those who they see as unrightful claimants. As a multi-million dollar industry, it is a profitable war, with companies like GPTZero, Copyleaks, Scribbr, QuillBot, TurnItIn, etc. gladly offering their mercenary services. These services are tainted by false positives, which result in harmful and disruptive consequences for students accused of using AI – *"classrooms remain plagued by anxiety and paranoia over the possibility of false accusations"* [18]. Such products result in meaningless squandering of student energy: *"[One student] became obsessive about avoiding another accusation. She screen-recorded herself on her laptop doing writing assignments. She worked in Google Docs to track her changes and create a digital paper trail. She even tried to tweak her vocabulary and syntax. [... Another student] says the majority of the time it takes him to complete an assignment is now spent tweaking wordings so he isn't falsely flagged"* [18]. Students turn against each other, passing the work of their peers through AI detection systems in acts of self-preservation [17]. So AI detectors position themselves as a mere decision support tool for teachers: *"nothing is 100% [...] it should be used to identify trends in students' work [...] kind of like a yellow flag for them to look into and use as an opportunity to speak to the students"*. Mercenaries are paid handsomely for their failures while teachers take the blame.

Such technical accessories to boundary work are not without controversy. At the heart of the controversy is a desire for secrecy, born of shame. For instance, OpenAI developed a watermarking tool that would enable the precise and accurate detection of ChatGPT-generated text, but demurred from releasing it, because despite one

2023 survey showing that *"people worldwide supported the idea of an AI detection tool by a margin of four to one"*, another survey of ChatGPT users revealed that *"69% believe cheating detection technology would lead to false accusations of using AI. Nearly 30% said they would use ChatGPT less if it deployed watermarks and a rival didn't"* [62]. Users clearly value the ability to use ChatGPT undetected. Without a shaming culture, this valorisation could not exist. Moreover, a majority of users are already concerned about false accusations of using AI. Without a shaming culture, this concern could not exist either.

These proprietary internal surveys (Seetharaman and Barnum [62]) showcase an asymmetry between when people are surveyed *in general* versus when people are surveyed *as users* of AI. If these are the same or an overlapping population (the details are not available), this asymmetry may be evidence of a revealed preference, which will be important to investigate in future work. Similarly, studies have found an asymmetric attitude to AI in personal communication: people are permissive and accepting of their *own* use of AI to draft outgoing or summarise incoming communications, but not of *others'* AI use. Consequently, they support automatic disclosure of others' AI use to themselves, but resist disclosure of their own AI use to others.

Zhang et al. [71] analyse the secret use of large language models, distinguishing two types of secrecy: passive non-disclosure and active concealment. Their survey finds several categories of tasks where participants tend to conceal their use of LLMs (e.g., sensitive topics, social communication, work tasks, school work, etc.). Importantly, they find that the *reasons* for concealment revolve entirely around the desire to avoid negative value judgements associated with the use of AI, both internal (e.g., questioning own competence, moral doubts, etc.) and external (e.g., fear of capability being critiqued, sincerity concerns in personal relationships, etc.). An internalised phobia, a self-shaming, is associated with AI use. So the gargoyle Shame squats atop the quivering conscience.

It is worth acknowledging that new media are often subject to shaming practices as a form of boundary work. Socrates (as told by Plato) shamed writing [51], Trithemius shamed the printed book [68],[2] Postman shamed the television [53], Carr shamed the Internet [13], all arguing – albeit with good reason – that those respective media result in the deterioration of mental faculties (a knowledge work degradation thesis, if you will). Similarly, arguments from creativity and arguments from technical basis have been deployed to question the status of DJing as a musical practice [58], the legitimacy of the camera obscura and camera lucida as tools for painters [42], and the status of photography as an art form [64].

Photography is a good model for AI shaming, as it appeared to directly antagonise a very specific kind of knowledge work professional: the painter. Hertzmann [28] provides an insightful analysis (which can be complemented by the opinionated but no less insightful account by Sontag [64]) of the antagonism between painting and photography. Antagonism was only one of three contemporary stances of the painting establishment towards photography in its early decades (the others being limited acceptance and full embrace), but it is the one that teaches us the most, because from

---

[2]From informal sources, I gather that the Neo-Confucian scholar and pedagogue Zhu Xi had raised similar objections to printed books some centuries earlier, but regrettably have been unable to trace an authoritative source to support this claim.

it we inherit a wealth of documented shaming practices that show AI shaming to be the reincarnation of an ancestral tendency. I will include only this delightful snipe from Baudelaire: *"If photography is allowed to supplement art in some of its functions, it will soon supplant or corrupt it altogether, thanks to the stupidity of the multitude which is its natural ally."*

4.1 AI Shaming Protects Class Boundaries

> All professions are conspiracies against the laity.
>
> Bernard Shaw, *The Doctor's Dilemma* (1906)

We have seen how AI shaming is a kind of boundary work practiced by knowledge workers. A key element of knowledge workers is their class identity. A brief examination of how AI appears to be affecting the class identity of knowledge workers is necessary to reveal the classism inherent in AI shaming.

The notion of "class" is heavily overloaded. Here I am referring to class as a socioeconomic identity derived from income, educational attainment, and occupation, following Kouaho and Epstein [37]. As a finer granule of class I will focus on *profession* (i.e., occupation), as members of a profession are typically (though not always) also homogeneous with respect to income and educational attainment.[3]

Boucher et al. [9] interview early-career game developers about Generative AI, finding that they were *"developing a new professional culture both* with *and* against *generative AI."* While they find that *"the most strongly expressed resistance to GAI came from artists who had ethical concerns about topics such as copyright and art theft"*, they also note that *"The artists in our study, who might have a strong sense of the distinctiveness of their artistic taste, tend to consider the broader community of artists in the game industry as their reference point when they discuss the implications of GAI on their work. Unlike the 'methodological individualism' of future predictions of economic impact, [they] see their future connected to others in the industry."* Similarly, Panchanadikar and Freeman [49] find an anxiety around indie game developers voiced as compassion for minority subprofessions of the industry (artists and voice actors) viewed as culturally valuable to the integrity of the profession as a whole.

As studies such as those by Boucher et al. [9] and Panchanadikar and Freeman [49] show, resistance to AI is grounded not in an individual, but collective apprehension of loss. The value of a profession is protected by the collective cooperation of all members in performing boundary work, in staging resistance, in casting shame, in maintaining class solidarity.

The basis for concern is legitimate. Economic analyses of the differential impact of AI on various occupations find a threat to knowledge work. For instance, Eloundou et al. [20] find that *"occupations needing "extensive preparation" (e.g., lawyers, pharmacists, database administrators) are more exposed than those with lower entry barriers (e.g., dishwashers, floor sanders). [...] higher-wage occupations are more exposed to LLMs than lower-wage occupations [...] The two job groups (clusters) that are most exposed to LLMs are "Scientists and Researchers," then "Technologists," such as software*

engineers and data scientists". We must be careful to note that "exposure" here means something closer to "potential surface area for application", and does not entail automation. The implied threat is not necessarily a loss of jobs, but rather of a carefully guarded moat ("extensive preparation", etc.) around such jobs.

Who crosses the moat? MIT economist David Autor is cautiously sanguine [7], claiming that AI *"can enable a larger set of workers equipped with necessary foundational training to perform higher-stakes decision-making tasks currently arrogated to elite experts, such as doctors, lawyers, software engineers and college professors [... thus] restoring the middle-skill, middle-class heart of the U.S. labor market that has been hollowed out by automation and globalization."* Shaming is part of the response of the professional (other parts include regulations and institutional codes) to extend this arrogation.

Johnson, in *Professions and Power* [32], finds that power in professions is predicated on a separation between producer (or service provider) and consumer, which creates a relationship of dependency (similar analyses of how occupational specialisation leads to a producer-consumer dichotomy and the deterioration of self-sufficiency are given by McLuhan [46] and Illich [30]). Moreover, power is derived from direct access to information processing. AI both frustrates the producer/consumer dichotomy and intermediates access to information processing, thus reducing professional power. In response, through shaming, professionals direct their ire at those they see as pretenders. Doctors have always derided home remedies, scientists have derided lay theories, sacerdotal colleges have derided folk mythologies and cosmogonies as heresy – the ability of individuals to "produce" their own healing, their own knowledge, their own salvation. These threaten professional power by undermining the relationship of dependency.[4]

4.2 Hard Boundaries on AI Use: Institutional Guidelines and Knowledge Materialism

Shaming as a phenomenon of informal social discourse crystallises into, and is in turn supported by, more hardened boundaries. Of relevance to our research community are institutional guidelines, rules, and codes of ethics around AI use, such as university regulations and those of publishers.

The Joint Council for Qualifications, a British umbrella organisation whose members administer among other things the most common high school qualifications in the UK, asks teachers to ensure that *"the final product is in [students'] own words, and isn't copied or paraphrased from another source such as an AI tool, and that the content reflects [students'] own independent work"* [3]. But as an example of AI misuse, this policy offers *"Copying or paraphrasing sections of AI-generated content so that the work submitted for assessment is no longer the student's own"*. The Marxian/Lockean invocations of "ownership" over work are repeated, the definition of AI misuse curiously circular. Closer to home, our publisher ACM states that AI use is permitted, as long as *"the resulting Work in its totality is an accurate representation of the authors' underlying work and novel intellectual contributions and is not primarily the result of the tool's generative capabilities"* [1]. One institution I am affiliated with states that *"A student using any unacknowledged content*

---

[3]There are cultural variations and nuances that this ignores; for instance in England, class identity is strongly hereditary – wealthy white collar workers may retain a "working class" identity if they were raised by working class parents [43], and there is no route by merit into the aristocracy, only by blood and marriage.

[4]More could be said about how this process is an accomplice to colonialism (e.g., see Freire [22] or Smith [63]) but is out of scope.

generated by artificial intelligence within a summative assessment as though it is their own work constitutes academic misconduct [...]" [4]. Another states that *"It is not acceptable to use GenAI tools to write your entire assessment and present this as your own work"* [2].

In the motivation for such guidelines, there is a recurring preoccupation with protecting "integrity" (scantily defined) through disclosure, monitoring, and oversight – some might say surveillance – of AI use [3, 4, 26]. Yet in practical terms, the protection of integrity also clearly entails protecting a notion of *ownership*. In other words, the morality of integrity and academic misconduct, its rhetoric of "unfair advantage" [2, 4], is ultimately derived from a rather *material* conceptualisation of knowledge work.

This conceptualisation of knowledge work as possessable material, with exchange value that one might be unfairly cheated out of, we might call *knowledge materialism*. Knowledge materialism is not newly precipitated by Generative AI, merely newly thrown into crisis by it. While a detailed discussion of the development of knowledge materialism is out of scope, it suffices here to note that education and the academy more broadly have long been undergoing a process of commodification and transactionalisation, a process that motivates many student and scholarly activities categorised as malpractice or cheating (e.g., as documented by Miles et al. [47]), as education and scholarship are pressed into the service of a hypercapitalised knowledge economy in which credentials and metrics increasingly translate into material prosperity.

However, to reconcile this materialism with the fundamental immateriality, disembodiment, and abstraction of knowledge work (not to mention a deeply ancestral scholastic revulsion for association with earthy, vulgar material concerns) has required the construction of an elaborate morality of knowledge work. Institutional guidelines demarcate not merely what is permitted, but also what is *virtuous*. The oblique burden of the academy to arbitrate morality in service of consolidating its material interests is quite analogous to the burden of the state to arbitrate creativity through intellectual property laws in service of maximally extracting value from its knowledge economy [58].

Just as the dual agenda of intellectual property laws in defining creativity while optimising innovation results in contradictions [58], so the dual agenda of institutional guidelines in defining morality while optimising commercial value results in contradictions in the internal logics for "appropriate" AI use. One guide for AI use in academic literature reviews opens with the promise that *"instead of discussing what Gen.AI can and cannot do, we discuss what we should allow Gen.AI to do, irrespective of its capabilities"* [67]. Yet after some discussion, it is abandoned: *"Do we prioritize the contribution of knowledge, irrespective of its source, or do we value maintaining control over the process of knowledge generation? These profound questions extend beyond the scope and intent of our current study."*

The ineluctable contradictions of ideological-material dual agendas also explain, in part, the challenge of developing workplace policies on Generative AI use. Unlike the academy, non-academic workplaces cannot from the concrete basis of knowledge materialism straightforwardly derive an abstract framework of morality. After all, the material exchange of knowledge work is the transparent motive of the for-profit enterprise, not a revolting encumbrance. Guidelines and boundaries on AI use mandated in non-academic workplaces are thus largely draped in the garb not of morality, but

of security concerns. Yet they cannot, until the latent materialism that underpins professional power is confronted, escape the inevitable layering and projection of morality onto such guidelines. Workers feel *"embarrassed [...] to admit to using shortcuts"* even if AI use is not explicitly banned, and employers enforcing explicit workplace AI bans insist that they *"don't want [employees] to think they're in trouble"*, yet refer to repeated use of AI in the morally charged vocabulary of *"recidivism"* [65].

Of course, such attitudes may in part be attributed to the enculturation of years of training and education during which knowledge materialism is explicitly moralised; the academy plays its role in fostering a thorough interiorisation of norms around shame and duty associated with class and profession that serve to condition individuals for participation in a society whose structures of material exchange depend on such norms being easy or free to enforce. This merely continues ancient strategies of societies of control that begin, not after the second world war as Deleuze would have it [19], but perhaps as far back as ancient Vedic or Confucian societies which similarly bound class and profession firmly to morality, duty, honour, and shame – or perhaps even further.[5]

4.3 Sumptuary Laws and Protectionism: Alternative Metaphors for The Effects of Generative AI on Class Relations

The effect of Generative AI on knowledge work is often compared to the industrial revolution. The strength of this analogy is that it captures how both movements of mechanisation similarly transform the relationship between labour, intellect, and material production [58]. However, they have (or are having) rather different effects on class relations. The industrial revolution did disenfranchise a similarly powerful and moated middle class of skilled craftsmen, but it was not without precedent in doing so; the 17th century had already seen the retreat of the medieval guild in response to the pressures of market economics and the invention of modern corporations. Rather, the defining characteristic of the industrial revolution's class relations was its introduction of the sclerotic and oppressive bourgeoisie-proletariat dyad. In contrast, economic analyses show a fundamentally different effect for Generative AI, which is, as has been abundantly mentioned, the opening of new routes into the middle class [7, 20].

Thus, the industrial revolution is an attractive but inadequate metaphor for Generative AI. Restrictions on AI use in modes that preserve existing notions of ownership over knowledge work, and AI shaming more broadly, bear more of a resemblance to *sumptuary laws* and *protectionist policies* of industry and labour. These restrictions recur across eras and geographies as formalisations of social boundaries threatened by technological innovations and economic circumstances. As a selection of examples one might

---

[5]There are clear distinctions between these societies and our latter-day societies of control as characterised by Deleuze – they are inhabited by whole individuals as opposed to atomised "dividuals", computational and numeric machineries of control where present are incomparable to those of present-day scale – but they share essential characteristic that the tightly boundaried and institutionalised forms of control in Foucault's disciplinary societies are rendered unnecessary by forms of control that are continuous and deeply interiorised to the point of being unconscious. Above all, they do not stage the clear oppression/freedom binary as do disciplinary societies.

consider: ancient Spartan and Roman laws regulating housing, currency, and dress, Edward III of England's attempts to ban merchants and servants from eating too much meat and the regulation of the pointiness of shoes by rank, the English ordinance of labourers limiting peasant mobility following the Black Death, Henry VIII of England's prohibition on crimson and blue velvet to be worn by anyone of lower rank than a knight, Louis XIV of France's Edict of Fontainebleau attempting to limit the exodus of skilled Huguenots, various attempts across Europe to regulate clothing according to social status between the 14th and 18th centuries [55], British textile laws in the 17th and 18th centuries that protected its inferior cotton textiles from the far superior cloth from Bengal, the USA's Chinese exclusion act, and so on.

One distinction worth making is between the types of regulations discussed above and the related notions of fashions and etiquette. Privileged classes have always developed norms (of speech, of behaviour, of dress, of music, of dance, and so on) to distinguish themselves and demonstrate "taste" [29]. Often these require great investments of time and resources to learn; displaying etiquette is thus a way to "conspicuously consume" such resources, the same principle underlying luxury goods [69]. The key difference between luxuries and the targets of sumptuary laws is that luxuries attempt to heighten and reposition the material barriers to class incursion, whereas sumptuary laws attempt merely to enforce them by fiat. Sumptuary laws are a recognition that the material barriers no longer exist and that the outgroup has access to the same resources; they are a last line of control. Yet their *prima facie* justification, much like AI shaming, is always something else: a question of taste, propriety, morality, ownership, degradation, and so on.

Thus, each case where laws are introduced to curtail social mobility that is newly enabled through technological innovation offers a potential alternative metaphor to the industrial revolution. Investigating these cases and the implications of their respective metaphors is left as an exercise for future research. It will be important to focus on responses to real threats to social boundaries due to shifts in the nature of material production and exchange, rather than merely perceived threats, or threats conjured for political expediency, as is so often the case with protectionism.

5 Societal Harms of AI Shaming

Caporusso [12] posits that Generative AI creates a mental condition described as *"creative displacement anxiety"* (CDA). CDA affects both producers and consumers of AI artefacts. Several distinct experiences that could induce CDA are described, including those that overlap with arguments in support of AI shaming, suggesting that performing this boundary work might be an auto-analgesic response to the experience of CDA, while at the same time increasing the cultural propensity for individuals to experience CDA by amplifying these anxieties across public discourse. At least one concrete CDA experience is exacerbated by (if not the product of) the culture of AI shaming: *"Imposter syndrome: artists using generative AI tools might feel they are not genuinely creating, leading to feelings of fraudulence"*. Kobiella et al. [36] document a similar phenomenon: *"[Participants experienced ...] a feeling of inadequacy, where participants believed their ideas couldn't compete with the AI-generated content: "If the machine is as good as me, then what use am I?""*

Kadoma et al. [33] investigate what they refer to as perceptual harm, defined as *"the harm caused to users when others perceive or suspect them of using AI"*. They find that *"people associate AI-stylized writing with lower quality"*, and that *"different groups are suspected of AI use at varying rates"*. Men are more likely to be suspected of AI use than women. Profiles *"suggestive of East Asian identity"* are more likely to be suspected of AI use than White Americans. The authors argue that suspicion of AI use (corresponding to "allegation" in our taxonomy) creates real harm on the basis that participants in their experiment were significantly less likely to "hire" freelancers that were suspected of AI use.

AI shaming may thus contribute to the racialisation of AI narratives. McInerney's concept of "Yellow Techno-Peril" illustrates how AI-related anxieties are racialised, as US media coverage of the so-called AI arms race with China *"casts the AI arms race as a civilisational conflict"* and revives the anti-Asian tropes of techno-Orientalism and the Yellow Peril [45]. In McInerney's analysis, the rhetoric of the AI arms race conflates AI advancement with racialised fears of Chinese dominance, reinforcing harmful stereotypes that Chinese individuals are untrustworthy and technologically threatening. This affects individual careers and reputations, and embeds xenophobic sentiment in AI discourse. The role of AI shaming in such discourse warrants further investigation.

It is clear that the unconfirmed suspicion of AI use can serve as an extension of (and veneer for) a pre-existing prejudice. Suspicion can only be prejudice because, in fact, humans are pretty poor at detecting AI-generated content, because we deploy flawed heuristics in attempting to do so [31], because we may in some cases even prefer AI-generated to human-generated content, and because we hypocritically reverse our judgement of an ostensibly human-generated artefact when we are told it is AI-generated and vice versa [52]. There is some evidence that expert AI users may become skilled at detecting AI-generated content [56], but since the scope of AI-generated content is so broad and volatile, and human communities so diverse, it seems likely that AI suspicion will continue to be dominated by pre-existing biases.

AI shaming can lead to negative consequences for shamer and shamed alike. For instance, a clothing merchandise company alleged that one of its contracted artists was using AI [5], resulting in reputational damage to the artist. When the company was in turn accused of sponsoring a witch hunt, they conducted an "investigation" in which they requested evidence of manual (i.e., not AI-assisted) work. After receiving satisfactory evidence (albeit still disputed by some commentators), the company issued an apology and compensated the artist. Even so, the apology demonstrates a continued value judgement associated with mechanical means of artistry, such as tracing, e.g., *"Although [the artist] admits to tracing some works in the past, we do not think that such a minor sin warrants the destruction of his reputation or the immense stress that we put him through"*, and the insistence that AI assistance renders artistic work valueless, e.g., *"I am much happier having to pay him knowing that he's legit, rather than not paying him knowing that he isn't."* The characterisation of tracing as "sin" recalls historical episodes that similarly vilified the camera lucida and photography [42, 64].

An analysis of 14 million paper abstracts estimates that approximately 10% of abstracts written in the first half of 2024 were processed using LLMs, and identifies tell-tale vocabulary whose use has

soared due to the introduction of LLM writing assistants, presided over by the notorious "delve" [35]. After reading this paper, my own attitude to this AI-shibboleth-in-chief has changed substantially. Delve has never been a star feature of my idiolect. Even so, I now consciously avoid it. If over the course of writing, my subconscious offers the 'D' word or one of its confederates, I bat it away with distaste. I might dive, explore, investigate, probe, examine, scrutinise, or interrogate, but never *delve* – the prospect of association with ChatGPT is too mortifying. Moloch has eaten the word and robbed its innocence. When I see it in others' writing, my suspicion is raised. Already the culture of shame has exerted a chilling effect on this author's vocabulary.

Alas, the story continues. Where did ChatGPT acquire its predilection for delving? From its data annotators, of course [27]. It turns out that "delve" is a common feature of Nigerian English. Nigerian commenters on Paul Graham's claim that *"delve [...] is a sign that the text was written by ChatGPT"* accused him of being *"blinded by the ripple effects of years of colonialism"*, with one commenter writing: *"Imagine after being force-fed colonial languages, being forced to speak it better than its owners then being told that no one used basic words like 'delve' in real life. Habibi, come to Nigeria"* [6]. To escape the mortifying pan of AI shame, we have leapt unwittingly into the mortifying fire of racism.

Already the cracks in the moral ideology of AI shaming, its boundary work, and its regulatory apparatuses are beginning to show. Liang et al. [41] have shown that the AI detectors discussed in Section 4 exhibit a significant bias against non-native English speakers, misclassifying over 60% of essays written by non-native English speakers versus 5% for native English speakers. Their chilling conclusion is that use of AI detectors is *"paving the way for undue harassment of specific non-native communities [... and can] restrict the visibility of non-native communities, potentially silencing diverse perspectives [...] penalize researchers from non-English-speaking countries [... and make] non-native students bear more risks of false accusations of cheating"*.

No less disquieting is a recent study of Black high school students documenting how they are disproportionately affected by AI shaming and regulations [66]. One student relates: *"It was kind of frowned upon, because it's used for cheating. So I never really used it. And in that way, I kind of frowned upon myself [...] as my teachers were telling me 'oh, it's a cheating device', I was realizing that it's also kind of racist."* In another's experience, *"Having to navigate racialized expectations for Black girls to do "twice the work for half the credit" while simultaneously navigating negative stereotypes about their inherent criminality not only made [the student] feel unsupported and "disposable," but it also made her uneasy about using AI to support her learning. Even though her AP language course encouraged students to use AI to help with second language acquisition, [the student] was reluctant to use the feature because of fears she would be falsely accused of cheating."*

Existing frameworks of epistemic injustice in Generative AI do account for some phenomena adjacent to AI shaming. Kay et al. [34] define "manipulative testimonial injustice", a class of injustices that includes *"the false accusation of deepfakes. This tactic exploits the increasing uncertainty surrounding the authenticity of digital media, creating a "liar's dividend" where even genuine evidence can*

be dismissed as fabricated [...] This weaponization of doubt and uncertainty further undermines the ability of marginalized groups to have their voices heard and their experiences validated." However, though superficially similar to slur and disparagement, the narrow intention of testimony manipulation is to undermine authority, which is only one of a number of possible intended outcomes of a slur. The broader set of AI shaming phenomena seem therefore to be under-explored in such frameworks, and present an opportunity to extend them.

6 Implications and Non-Implications

Resisting AI shaming. As may be evident from the foregoing arc of argumentation, this paper takes the position that AI shaming is undesirable and harmful. The most immediate remedial measures can therefore be taken by our own research community, by not participating in AI shaming and by resisting it when it arises in our professional and institutional discourse. The limited scope of this paper does not permit expansion into strategies for policymakers or AI developers, or proposals for educational reforms. However, the catalogue of harms in Section 5 yields a modest starting point; we might consider how each individual harm might be tackled. For instance, creative displacement anxiety around art that involves AI use might be mitigated by introducing exhibitions, awards, and galleries that celebrate the human ingenuity involved in the artful use of AI (though the definition of "artful" here might itself enable new forms of solidarity and exclusion, cf. the *Salon des Refusés*), or by advocating for degree programs that incorporate judicious and principled AI use, to dispel the generalised and vague stigma surrounding AI use among staff and students. This is hardly the first time such measures have been proposed, but it is perhaps the first time they have been proposed specifically as a countervailing manoeuvre against AI shaming culture.

Does AI erase or merely rearrange class boundaries? The evolving nature of AI shaming raises further questions about its long-term implications for class structures in knowledge work. The erosion of traditional boundaries in knowledge work might shift the burden of maintaining professional standards from institutions to individuals. If there is no longer a broadly accepted framework for evaluating expertise, workers may find themselves responsible for continuously justifying and demonstrating their own competence. This could privilege those who possess a certain combination of technological aptitude and entrepreneurial character (as mentioned in Section 2) while disadvantaging others. The result may not be the dissolution of hierarchies, but rather their reconfiguration into new hierarchies that favour those with the skills, disposition, and resources to thrive in what might be called a post-epistocratic knowledge work landscape. This raises the broader question of whether shame itself may reconfigure along new lines, or whether a world without AI shaming is possible. And if the latter is possible, whether it would be one without any stratification, or whether the mechanisms of exclusion would simply reformulate themselves. A critical next step in understanding AI shaming, therefore, is to explore how AI-driven changes in knowledge work intersect with broader structures of power in labour. These questions – with answers that are as volatile as the sociotechnical landscape they query – require careful, continuous, and multidisciplinary analysis.

Distinguishing justified critique from classism. A more vexing challenge is to balance critique of AI shaming with legitimate critiques of AI development. As mentioned in Section 3, arguments marshalled in support of AI shaming are not invalid.[6] An accusation of classism, such as this paper offers, can easily be weaponised to dismiss very real issues such as work degradation and the effects of "slop" on the information ecosystem. Critiques of the power dynamics of AI, particularly those taking aim at the asymmetry between system developers and users, observe that shaming and shaming-adjacent behaviour may be a mode of retaliative expression against powerful firms. It may be the only such mode available to those whose lives and labour are affected by the practices that enable the creation of Generative AI, and the consequences that follow its deployment.

To level the charge of classism against such critiques would be a gross misinterpretation of the intent of this paper, which is to visibilise and contest AI shaming *in its specific capacity as class boundary work*, not to disqualify real objections to the technology and its power dynamics or disenfranchise those for whom AI shaming is purely a mode of counterhegemonic resistance.

The spectrum of possible stances on AI shaming risks devolving into a false dichotomy: either AI shaming is permissible because it enables legitimate critiques of AI, or it is not permissible because it reinforces classist boundary work. For at least two reasons, this dichotomy is a false one. First, the classic populist misdirection play of the elite is to convince the oppressed classes to channel their pain and frustration downwards, into resentment and anger towards those even more oppressed by them. Consequently, both of the following can occur simultaneously: system developers exercising power over knowledge workers, and knowledge workers exercising power over less privileged "outsider" classes. Second, this power dynamic is evolving. It is reasonable to optimistically characterise Generative AI in particular as a technology where users are not as disenfranchised as they are commonly assumed to be. The economy of Generative AI is far from a state of monopoly or oligopoly. Advances in model training and miniaturisation, open datasets and weights, the existence of a vibrant, diverse, and competitive market landscape, etc., all indicate that the commoditisation of language models *per se* is well underway. Producers have far less power in commoditised markets than do consumers and regulators.

AI shaming conflates boundary work with justifiable objections. What stance, then, is appropriate? Can critiques of AI development couched in shaming be separated from harmful classism? Developing a principled answer to this question is beyond the scope of this paper, however, the desired outcome is clearly to ensure that any genuine critiques embedded in shaming discourse are directed at the right people, for the right reasons, in the right form.

7 Conclusion

This paper opens a critical discussion of social practices that associate negative judgements with the use of AI, under the general term "AI shaming". It has discussed how AI shaming can manifest in many forms, depending on whether the use or non-use of AI

is definitively established, and whether a person (or group of people) or the artefact (or set of artefacts) is the target of the negative judgement. It has further discussed the common arguments cited to support shaming, including that AI harms people, AI cannot be creative, and that AI output is qualitatively poor.

An alternative account of AI shaming is then presented: that AI shaming is a form of boundary work that knowledge workers undertake to demarcate themselves. In doing so, they protect class boundaries that are perceived as being eroded. Beyond shaming, this boundary work is also embedded into AI detection tools, institutional regulations, and codes of ethics around AI use. It is suggested that rather than the industrial revolution, we might look at sumptuary laws and protectionist policies as alternative metaphors for the effects of Generative AI on society.

Finally, we see how AI shaming results in systematic harms, especially to racial minorities and non-native English speakers by amplifying discrimination, but also in society more broadly by exerting chilling effects and inducing psychological anxiety. It is therefore incumbent on us not to participate in AI shaming, to resist it when we observe others doing it, and to re-examine our institutional regulations and ensure that we are not merely morality-washing lipstick onto a solidarity pig.

Acknowledgments

References

[1] 2024. ACM Policy on Authorship. https://www.acm.org/publications/policies/frequently-asked-questions Accessed: 2024-12-02.

[2] 2024. Engaging with Generative AI in Your Education and Assessment. https://www.ucl.ac.uk/students/exams-and-assessments/assessment-success-guide/engaging-generative-ai-your-education-and-assessment Accessed: 2024-12-02..

[3] 2024. *JCQ AI Use in Assessments: Protecting the Integrity of Qualifications Guidance.* Technical Report. Joint Council for Qualifications. https://www.jcq.org.uk/wp-content/uploads/2024/07/AI-Use-in-Assessments_Feb24_v6.pdf Group report with no individual authors listed..

[4] 2024. Plagiarism and Academic Misconduct: Artificial Intelligence. https://www.plagiarism.admin.cam.ac.uk/what-academic-misconduct/artificial-intelligence Accessed: 2024-12-02..

[5] @IcedTeaClothing (Iced Tea Aesthetics). 2024. *We accused an artist of selling us traced AI. Accusations, apology compensation & fundraiser.* https://x.com/IcedTeaClothing/status/1834677501148512763 Twitter (X) post, Accessed: 2024-11-20.

[6] Eze Anaba (Ed.). 2024. Nigerians Tackle American Author for Claiming 'Delve' Is Only Used by ChatGPT. *Vanguard* (10 April 2024). https://www.vanguardngr.com/2024/04/nigerians-tackle-american-author-for-claiming-delve-is-only-used-by-chatgpt/ Accessed: 2024-12-02..

[7] David Autor. 2024. How AI Could Help Rebuild the Middle Class. *Noēma Magazine* (12 February 2024). https://www.noemamag.com/how-ai-could-help-rebuild-the-middle-class/ Accessed: 2024-11-26.

[8] Emily Bender. 2024. *Information Literacy and Chatbots as Search.* https://buttondown.com/maiht3k/archive/information-literacy-and-chatbots-as-search/ Accessed: 2024-11-20.

[9] Josiah D Boucher, Gillian Smith, and Yunus Doğan Telliel. 2024. Is Resistance Futile?: Early Career Game Developers, Generative AI, and Ethical Skepticism. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–13.

[10] Harry Braverman. 1974. *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century.* Monthly Review Press, United States.

[11] Blake Brittain. 2024. Artist Sues After US Rejects Copyright for AI-Generated Image. *Reuters* (September 2024). https://www.reuters.com/legal/litigation/artist-sues-after-us-rejects-copyright-ai-generated-image-2024-09-26/ Accessed: 2024-11-20.

---

[6]Indeed, some of my own work explores an aspect of work degradation, namely, the deterioration in critical thinking that may ensue from the unreflective application of Generative AI in knowledge work [40, 59–61]. I have also aimed to draw attention to the socially problematic nature of the "human-AI collaboration" metaphor [57].

[12] Nicholas Caporusso. 2023. Generative artificial intelligence and the emergence of creative displacement anxiety. *Research Directs in Psychology and Behavior* 3, 1 (2023).

[13] Nicholas G. Carr. 2010. *The Shallows: What the Internet Is Doing to Our Brains.* W. W. Norton & Company. 280 pages.

[14] Nick Cave. 2023. *ChatGPT: The Soul Eater.* https://www.theredhandfiles.com/chatgpt-making-things-faster-and-easier/ Issue #248, The Red Hand Files. Accessed: 2024-11-19.

[15] Kyle Chayka. 2024. How to Opt Out of A.I. Online. *The New Yorker* (October 2024). https://www.newyorker.com/culture/infinite-scroll/how-to-opt-out-of-ai-online Accessed: 2024-11-14.

[16] Ted Chiang. 2024. Why A.I. Isn't Going to Make Art. *The New Yorker* (August 2024). https://www.newyorker.com/culture/the-weekend-essay/why-ai-isnt-going-to-make-art Accessed: 2024-11-14.

[17] Will Coldwell. 2024. 'I received a first but it felt tainted and undeserved': inside the university AI cheating crisis. *The Guardian* (15 Dec. 2024). https://amp.theguardian.com/technology/2024/dec/15/i-received-a-first-but-it-felt-tainted-and-undeserved-inside-the-university-ai-cheating-crisis Published 07:00 GMT.

[18] Jackie Davalos and Leon Yin. 2024. AI Detectors Falsely Accuse Students of Cheating—With Big Consequences. *Bloomberg Businessweek* (October 2024). https://www.bloomberg.com/news/features/2024-10-18/do-ai-detectors-work-students-face-false-cheating-accusations Accessed: 2024-11-20.

[19] Gilles Deleuze. 1992. Postscript on the Societies of Control. *October* 59 (1992), 3–7. http://www.jstor.org/stable/778828

[20] Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. 2024. GPTs are GPTs: Labor market impact potential of LLMs. *Science* 384, 6702 (2024), 1306–1308.

[21] Michel Foucault. 1977. *Discipline and Punish.* Pantheon Books, New York, NY. First English edition, translated from French (Surveiller et punir, Gallimard, 1975). 318 pages, ISBN: 0-394-49942-5.

[22] Paulo Freire. 1968. *Pedagogy of the Oppressed.* Brazil. Original title: *Pedagogia do Oprimido..*

[23] Helen Gaskell. 2024. *Do you know how LLMs work?* https://www.linkedin.com/feed/update/urn:li:activity:7252381573616926720/ LinkedIn post, Archive URL: https://archive.is/CDLDN, Accessed: 2024-11-20.

[24] Jack Apollo George. 2024. 'If Journalism Is Going Up in Smoke, I Might as Well Get High Off the Fumes': Confessions of a Chatbot Helper. *The Guardian* (September 2024). https://www.theguardian.com/technology/article/2024/sep/07/if-journalism-is-going-up-in-smoke-i-might-as-well-get-high-off-the-fumes-confessions-of-a-chatbot-helper Accessed: 2024-11-14.

[25] Thomas F Gieryn. 1983. Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American sociological review* (1983), 781–795.

[26] O. H. Halvorson. 2024. AI in academia: Policy development, ethics, and curriculum design. *School of Information Student Research Journal* 14, 1 (2024). https://scholarworks.sjsu.edu/ischoolsrj/vol14/iss1/1 Retrieved from ScholarWorks at SJSU.

[27] Alex Hern. 2024. TechScape: How Cheap, Outsourced Labour in Africa is Shaping AI English. *The Guardian: TechScape Newsletter* (16 April 2024). https://www.theguardian.com/technology/2024/apr/16/techscape-how-cheap-outsourced-labour-in-africa-is-shaping-ai-english Accessed: 2024-12-02..

[28] Aaron Hertzmann. 2018. Can computers create art?. In *Arts*, Vol. 7. MDPI, 18.

[29] Samuel Hughes. 2024. Cheap Ornament and Status Games: Was Modernism Originally a Way to Signal Taste Instead of Wealth? *Works in Progress Newsletter* (15 October 2024). https://www.worksinprogress.news/p/cheap-ornament-and-status-games Accessed: 2024-12-02..

[30] Ivan Illich. 1981. *Shadow Work.* Marion Boyars.

[31] Maurice Jakesch, Jeffrey T. Hancock, and Mor Naaman. 2023. Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences* 120, 11 (2023), e2208839120. https://doi.org/10.1073/pnas.2208839120 arXiv:https://pnas.org/doi/pdf/10.1073/pnas.2208839120

[32] Terence J Johnson. 1972. *Professions and Power.* Studies in Sociology.

[33] Kowe Kadoma, Danaë Metaxa, and Mor Naaman. 2024. Generative AI and Perceptual Harms: Who's Suspected of using LLMs? arXiv:2410.00906 [cs.HC] https://arxiv.org/abs/2410.00906

[34] Jackie Kay, Atoosa Kasirzadeh, and Shakir Mohamed. 2024. Epistemic Injustice in Generative AI. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 7, 1 (Oct. 2024), 684–697. https://doi.org/10.1609/aies.v7i1.31671

[35] Dmitry Kobak, Rita González-Márquez, Emőke-Ágnes Horvát, and Jan Lause. 2024. Delving into ChatGPT usage in academic writing through excess vocabulary. *arXiv preprint arXiv:2406.07016* (2024).

[36] Charlotte Kobiella, Yarhy Said Flores López, Franz Waltenberger, Fiona Draxler, and Albrecht Schmidt. 2024. "If the Machine Is As Good As Me, Then What Use Am I?"–How the Use of ChatGPT Changes Young Professionals' Perception of Productivity and Accomplishment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–16.

[37] Whitney-Jocelyn Kouaho and Daniel A Epstein. 2024. Socioeconomic Class in Physical Activity Wearables Research and Design. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–15.

[38] Pia Lauritzen. 2024. Why Ben Affleck's Take On AI Is Only Half The Truth. *Forbes* (20 Nov. 2024). https://www.forbes.com/sites/pialauritzen/2024/11/20/why-ben-afflecks-take-on-ai-is-only-half-the-truth/

[39] Joe Lazer (Lazauskas). 2024. *The Human Internet Is Dying.* https://www.linkedin.com/posts/joe-lazer-lazauskas-8b442026_the-human-internet-is-dying-someone-just-activity-7252108012759195648-EkBs/ LinkedIn post, Archive URL: https://archive.is/hcQKj, Accessed: 2024-11-20.

[40] Hao-Ping (Hank) Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, USA, Yokohama, Japan, 23. https://doi.org/10.1145/3706598.3713778

[41] Weixin Liang, Mert Yuksekgonul, Yining Mao, Eric Wu, and James Zou. 2023. GPT detectors are biased against non-native English writers. *Patterns* 4, 7 (2023).

[42] Christoph Lüthy. 2005. Hockney's Secret Knowledge, Vanvitelli's Camera Obscura. *Early Science and Medicine* 10, 2 (2005), 315–339. http://www.jstor.org/stable/4130315

[43] Abigail Marks and Chris Baldry. 2009. Stuck in the middle with who? The class identity of knowledge workers. *Work, Employment and Society* 23, 1 (2009), 49–65.

[44] James Marriott. 2024. Why Human Creativity Matters in the Age of AI. *Engelsberg Ideas* (October 2024). https://engelsbergideas.com/essays/why-human-creativity-matters-in-the-age-of-ai/ Accessed: 2024-11-14.

[45] Kerry McInerney. 2024. Yellow Techno-Peril: The 'Clash of Civilizations' and anti-Chinese racial rhetoric in the US–China AI arms race. *Big Data & Society* 11, 2 (2024), 20539517241227873. https://doi.org/10.1177/20539517241227873 arXiv:https://doi.org/10.1177/20539517241227873

[46] Marshall McLuhan. 1964. *Understanding Media: The Extensions of Man.* McGraw-Hill, Canada. 318 pages.

[47] Paula J Miles, Martin Campbell, and Graeme D Ruxton. 2022. Why students cheat and how understanding this can help reduce the frequency of academic misconduct in higher education: A literature review. *Journal of Undergraduate Neuroscience Education* 20, 2 (2022), A150.

[48] Evelyn Van Orden. 2024. *LinkedIn post.* https://www.linkedin.com/posts/evelyn-van-orden-41644153_your-grandchildren-will-be-the-last-generation-activity-7257297931345539073-kk_P/ Accessed: 2024-11-14.

[49] Ruchi Panchanadikar and Guo Freeman. 2024. "I'm a Solo Developer but AI is My New Ill-Informed Co-Worker": Envisioning and Designing Generative AI to Support Indie Game Development. *Proceedings of the ACM on Human-Computer Interaction* 8, CHI PLAY (2024), 1–26.

[50] penguinz0. 2024. *Delusional AI "Artist".* https://www.youtube.com/watch?v=e3XRb-5qaQk YouTube video, Accessed: 2024-11-20.

[51] Plato. 1997. Phaedrus. In *Complete Works,* J. M. Cooper (Ed.). Hackett, Indianapolis, IN, 551–552.

[52] Brian Porter and Edouard Machery. 2024. AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably. *Scientific Reports* 14, 1 (2024), 26133.

[53] Neil Postman. 1985. *Amusing Ourselves to Death: Public Discourse in the Age of Show Business.* Viking Penguin, United States. 184 pages.

[54] Max Read. 2024. Drowning in Slop. *New York Magazine* (September 2024). https://nymag.com/intelligencer/article/ai-generated-content-internet-online-slop-spam.html Accessed: 2024-11-14.

[55] Giorgio Riello and Ulinka Rublack (Eds.). 2019. *The Right to Dress: Sumptuary Laws in a Global Perspective, c.1200–1800.* Cambridge University Press.

[56] Jenna Russell, Marzena Karpinska, and Mohit Iyyer. 2025. People who frequently use ChatGPT for writing tasks are accurate and robust detectors of AI-generated text. arXiv:2501.15654 [cs.CL] https://arxiv.org/abs/2501.15654

[57] Advait Sarkar. 2023. Enough With "Human-AI Collaboration". In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23).* Association for Computing Machinery, New York, NY, USA, Article 415, 8 pages. https://doi.org/10.1145/3544549.3582735

[58] Advait Sarkar. 2023. Exploring Perspectives on the Impact of Artificial Intelligence on the Creativity of Knowledge Work: Beyond Mechanised Plagiarism and Stochastic Parrots. In *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work* (Oldenburg, Germany) *(CHIWORK '23).* Association for Computing Machinery, New York, NY, USA, Article 13, 17 pages. https://doi.org/10.1145/3596671.3597650

[59] Advait Sarkar. 2024. AI Should Challenge, Not Obey. *Commun. ACM* (Sept. 2024), 5 pages. https://doi.org/10.1145/3649404 Online First.

[60] Advait Sarkar. 2024. Intention Is All You Need. In *Proceedings of the 35th Annual Conference of the Psychology of Programming Interest Group (PPIG 2024).*

[61] Advait Sarkar, Xiaotong (Tone) Xu, Neil Toronto, Ian Drosos, and Christian Poelitz. 2024. When Copilot Becomes Autopilot: Generative AI's Critical Risk to Knowledge Work and a Critical Solution. In *Proceedings of the Annual Conference*

of the European Spreadsheet Risks Interest Group (EuSpRIG 2024).

[62] Deepa Seetharaman and Matt Barnum. 2024. There's a Tool to Catch Students Cheating With ChatGPT. OpenAI Hasn't Released It. *The Wall Street Journal* (4 August 2024). https://www.wsj.com/tech/ai/openai-tool-chatgpt-cheating-writing-135b755a Archive URL: https://archive.is/BcVHA, Accessed: 2024-11-20.

[63] Linda Tuhiwai Smith. 1999. *Decolonizing Methodologies: Research and Indigenous Peoples.* Zed Books, New Zealand.

[64] Susan Sontag. 1977. *On Photography.* Farrar, Straus and Giroux, United States. 207 pages.

[65] Stephanie Stacey. 2024. Bosses struggle to police workers' use of AI. *Financial Times* (16 Dec. 2024). https://www.ft.com/content/cd08b45d-12dc-447e-bd59-1c366a7e6396

[66] T. C. Tanksley. 2024. "We're Changing the System with This One": Black Students Using Critical Race Algorithmic Literacies to Subvert and Survive AI-Mediated Racism in School. *English Teaching: Practice & Critique* 23, 1 (2024), 36–56. https://doi.org/10.1108/ETPC-08-2023-0102

[67] Fabian Tingelhoff, Micha Brugger, and Jan Marco Leimeister. 2024. A guide for structured literature reviews in business research: The state-of-the-art and how to integrate generative artificial intelligence. *Journal of Information Technology* (2024), 02683962241304105. https://doi.org/10.1177/02683962241304105 arXiv:https://doi.org/10.1177/02683962241304105

[68] Johannes Trithemius. 1494. *De Laude Scriptorum Manualium.* Originally printed in 1494. An English translation of the entire work is available in Behrendt (1974)..

[69] Thorstein Veblen. 1899. *The Theory of the Leisure Class: An Economic Study in the Evolution of Institutions.* Macmillan, United States. 400 pages.

[70] Fiona Warburton. 2024. From Gamification to Game-Based Learning. *JSTOR Daily* (July 2024). https://daily.jstor.org/from-gamification-to-game-based-learning/ Accessed: 2024-11-26.

[71] Zhiping Zhang, Chenxinran Shen, Bingsheng Yao, Dakuo Wang, and Tianshi Li. 2024. Secret Use of Large Language Model (LLM). arXiv:2409.19450 [cs.HC] https://arxiv.org/abs/2409.19450