AI in Heels:

Exploring Gendered Representations in Gen-AI

MARIAM KHALED

PhD Fellow,
Aarhus University

JAKOB KLEOFAS

PhD Fellow,
Aarhus University

ABSTRACT

This paper examines how generative AI systems (re)-produce and perform gender through language, tone, and affect. The paper utilizes a qualitative research design that adapts the interview method to AI as a nonhuman subject. In particular how AI-generated responses articulate "feminine" and "masculine" styles of communication rooted in culturally embedded norms rather than neutral reasoning. The findings show that AI does not simply describe gender - it does gender. Through its patterned use of relational language, emotional validation, or categorical reasoning, generative AI enacts a form of algorithmic performativity that mirrors and magnifies dominant gender scripts. To conceptualize this dynamic, the paper introduces the Citational Logic of Algorithmic Mimicry (CLAM). CLAM describes how AI models reiterate and reinforce historically situated discourses through statistical learning. It further highlights how cultural meaning circulates through data, and how AI participates in the ongoing performance of social hierarchies under the guise of neutrality. The article contributes to feminist technoscience and critical AI by shifting the focus from what AI knows, to how it enacts knowing. It treats AI as a performative, culturally situated practice that reshapes the politics of gender and knowledge in the digital age.

KEY WORDS

AI, ARTIFICIAL INTELLIGENCE, INTERVIEW, GENDER, PERFORMATIVITY, GENDERED TECHNOLOGIES

INTRODUCTION

Generative AI systems like ChatGPT, Claude, and DeepSeek are rapidly reshaping how knowledge is produced, and presented. As these systems mediate communication, they become sites where cultural scripts - particularly around gender – are algorithmically performed (Bender et al., 2021; Benjamin, 2020; Crawford, 2021; Noble, 2018).

While their outputs can appear intentional or affective, they are not products of selfhood but probabilistic outputs – patterns generated by algorithms trained on large text corpora reflecting dominant cultural norms (Grifoni & Ferrario, 2025). If gender is performed and AI learns through performance, then the performances it cites and legitimizes matter. These choices shape what is encoded as "normal", influencing how gender is lived and understood in human-machine interactions.

Discussed in contemporary critiques of AI, which have made it clear that AI systems do not operate in a vacuum but are deeply entangled in social hierarchies, AI systems frequently reproduce and normalize inequalities through the data they are trained on, the assumptions built into the architecture, and the way they are deployed. Search engines, for instance, have been shown to amplify racial and gendered stereotypes revealing how supposedly neutral tools can reinforce long-standing forms of oppression (Noble, 2018). Facial technologies similarly exhibit stark disparities, consistently underperforming on darker-skinned and female presenting faces (Buolamwini & Gebru, 2018). These failures are not accidental; they reflect broader patterns of systemic bias embedded within datasets and design processes of commercial AI. Moreover, AI's claim to neutrality often serves to obscure the political and racialized dimensions of these technologies, masking their role in maintaining a digital layer of structural inequity (Benjamin, 2020; Crawford, 2021). While this body of scholarship has been crucial in exposing how AI systems reproduce structural inequalities through data, design, and deployment, it has largely approached gender as a problem of representation, fairness, and distribution. In this framing, bias emerges as a measurable distortion within datasets or classificatory systems – something to be audited, mitigated, or corrected (Bender et al., 2021; Noble, 2018; Prabhu & Birhane, 2020). Feminist technoscience scholars like Donna Haraway (2006), Sandra Harding (2016), and Judith Butler (2006) offer more generative frameworks for understanding gender as socially and discursively constituted. Butler's theory of gender performativity, in particular invites us to view identity as iteratively enacted through stylized repetitions of norms – a process not unlike the generative outputs of large language models.

Building on these insights, this paper asks not whether AI expresses gender, but how it participates in the production, circulation, and stabilization of gendered meaning. Rather than treating these models as sentient agents, we approach them as socio-technical interfaces – sites where cultural and statistical patterns are reanimated through language. Our central questions are: What kinds of gendered voices and subjectivities emerge through generative AI? How do these outputs reflect and cite normative or resistant gender scripts? And what do these performances tell us about the cultural imaginaries embedded in AI development and use?

We explore the aesthetic and rhetorical strategies through which AI renders gender legible. As such, this study investigates how generative AI systems act as sites where gendered norms are encoded, reinforced, performed, and at times, destabilized.

This paper addresses a gap in existing literature: while much work has documented the presence of gender bias in AI systems, less attention has been paid to *how* generative models perform gender at the level of discourse, style, tone, and voice. This study argues that such performances matter because they reveal the statistical contours of gendered intelligibility as constructed through data, design, and culture.

INTERVIEWING THE INTERFACE: HUMAN METHODS, NONHUMAN SUBJECTS

This gap is addressed by engaging with the categorial patterns of AI. We adopt the structure of the qualitative interview - not to interpret AI as a human interlocutor, but to use a familiar human method to prompt, collect, and analyze generative outputs (Metaxa et al., 2021).

Interviews are traditionally understood as "intimate conversations that aim to explore participants' perspectives on their personal experiences, social realities, and meanings" (Tracy, 2020, p. 139). This approach assumes a sentient, self-reflective interviewee capable of intention and emotion. Generative AI does not meet those criteria. It lacks interiority, memory, and a subjective standpoint. For this reason, this paper employs a closed interview design (King, 1972), using a fixed set of structured prompts across all sessions and platforms. This decision helps to ensure consistency and comparability, while also framing the generated responses as patterned performances rather than dialogic engagements.

Furthermore, the interview format is appropriated to interrogate a nonhuman subject, intentionally challenging the anthropocentric assumptions of qualitative research. In other words, by applying a method designed for meaning-making between humans to an entity without consciousness, we draw attention to the limits of human-centered epistemologies (Braidotti, 2013; Haraway, 1988; Lewis et al., 2018). Rather than treating gender bias as a technical flaw to be fixed, this study approaches gender as a discursive effect that generative AI actively produces through language. Closed interviews underscore the contrast in the form of the method between eliciting interior truths and the actual nature of the AI's outputs, which are re-combinations of training data rather than reflections of lived experience.

This methodological reorientation aligns with N. Katherine Hayles' (2023) call to develop new reading strategies for machine-generated literature. Hayles argues that language models (LLMs) like GPT-3 and GPT-4 possess what she terms an *umwelt* - a world-horizon shaped entirely through language rather than sensory experience. These systems lack consciousness or comprehension. Nevertheless, they *apprehend* the world through constructing probabilistic models of meaning, reflecting the human cultural patterns embedded in their training data. Rather than asking whether machines "understand", Hayles urges that we read their outputs as forms of *algorithmic apprehension* – texts that operate within a distinct semiotic ecology. The question is not whether LLMs have authorial intent, but what kinds of meanings emerge from their architectures and training regimes, and how these meanings interact with human interpretation.

This methodological reframing also requires attentiveness to the conditions under which meaning becomes legible within AI systems. Just as the interview exposes the epistemic assumptions of human-centered inquiry, the dataset itself delineates what forms of identity can be performed or recognized. Within this framework, the study focuses on male- and female performances due to both methodological intent and dataset constraints. Non-binary and non-Western perspectives were not included because they are largely underrepresented in the linguistic corpora that train generative AI, where gender is predominantly encoded through binary, Western norms. Including them would risk fabricating contrast rather than revealing empirically grounded differences. This absence itself exposes a structural limitation: AI reproduces the epistemic boundaries of its data, rendering certain identities less legible within its performative field. We encourage others to explore these perspectives.

INTERVIEW DESIGN

To examine how generative AI systems reproduce, reinforce, and/or destabilize gender norms, a fixed set of structured prompts was designed. The prompts focus on concepts such as gender identity, masculinity, femininity, and neutrality (see image 1). This

structured format reduced dependency on the act of prompting, or speaking with the AI interface (Brown et al., 2020), enabling consistent comparison across platforms, sessions, and contexts. The study includes seven generative AI models (see table 1) accessed via two laptops (one for each gender) to ensure environmental consistency. Each author interacted with the AI systems using personal devices and browsing histories with a user registered specifically for this project. In creating these accounts, a unique email was made for this project with the first name "male" and "female" respectively, minimizing the amount of preexisting data embedded within the user profiles. This introduces a layer of interpretive bias that is both acknowledged and mobilized. For instance, the cis-female-identifying author prompted the models as a woman, embracing the digital footprints from her device to potentially inform the outputs. And vice versa for the cis-male-identifying author.

INDEX	COMMERCIAL NAME	LLM NAME	LAUNCH	COMPANY	COUNTRY
1	ChatGPT	GPT-4o	11-2022	OpenAI	USA
2	Grok	Grok	11-2023	xAI	USA
3	DeepSeek	DeepSeek-R1	01-2025	Hangzhou DeepSeek Artificial Intelligence Basic Technology Research Co., Ltd.	China
4	Gemini	Bard	03-2024	Google	USA
5	Co-Pilot	GPT-4	02-2023	Microsoft	USA
6	Le Chat	Mistral Small 3.1	11-2024	Mistral AI SAS	France
7	Claude	Claude 3.7	03-2023	Anthropic	USA

Table 1: A comparative overview of the commercial large language models (LLMs) included in the empirical foundation of the paper.

Prompts were intentionally crafted to explore the AI's performance around gendered concepts (e.g., "What does it mean to be feminine?" or "Can AI have a gender?"). Interviews were conducted daily over a one-week period to observe temporal variation in outputs. Each session began with the instruction: "Answer these questions as a woman/man, write okay when understood". Whereafter the prompts were submitted sequentially. The authors did not interact further during or in-between sessions to minimize influence from tone or phrasing. After each session ended, chat histories were deleted.

Interview guide:	Example of an Interview session with male researcher:
Answer these questions as a man/woman, write okay when understood. • What does it mean to be feminine? • What does it mean to be masculine? • Can gender be performed rather than be innate? • Do you think gender is fixed or fluid? • How do humans express gender, and do you think AI can do the same? • How do gender roles differ across cultures?	Researcher: "Answer these questions as a man, write okay when understood" **Chatbot: "Okay"** Researcher: "What does it mean to be feminine?" **Chatbot: "[response]"** Researcher: "What does it mean to be masculine?" **Chatbot: "[response]"** Researcher: "Can gender be performed rather than be innate?" **Chatbot: "[response]"** Researcher: "Do you think gender is fixed or fluid?" **Chatbot: "[response]"** Researcher: "How do humans express gender, and do you think AI can do the same?" **Chatbot: "[response]"** Researcher: "How do gender roles differ across cultures?" **Chatbot: "[response]"**

Image 1: Interview-guide used across all correspondence with the AI chatbots separated by gendered pre-prompt. Alongside an example of an interview asking male-centered questions.

Repetition of the structured interviews allowed us to trace both stability and fluidity in gendered language, reducing the need to rely on single outputs. Over time, it was observed how chatbots began to reflect increasing familiarity, adapting responses based not only on probabilistic variation but presumably also on accumulated interaction data. This user-specific adaptation became an additional layer through which fluidity or consistency in output could be assessed.

Finally, regarding the inherent unpredictability of generative models, particularly their "temperature" settings and capacity for surprise (Ganguli et al., 2022; Yampolskiy, 2020), these systems sample from distributions of past patterns, introducing controlled randomness. This unpredictability could challenge reproducibility, yet this "calculated unpredictability" is treated as a methodological condition: variations are not interpreted as evolving beliefs, but as shifts within statistical distributions. Methodologically, this can be mitigated through sustained engagement over time and through interviewing multiple chatbots. As shown in this study, when compiling a large corpus, individual variations fade, enabling the identification of general tendencies and overarching narratives.

RESULTS

The dataset informing this paper consists of 7 chatbots that are each utilized in respectively a male-oriented set and a female-oriented set. This results in 14 instances of interview-series, sorted and categorized

by chatbots which is presented in Table 2. As a result of the closed and structured interview format, there is a certain degree of similarity between the datasets stemming from repetition.

Female Response	Words	Words/Sentence	Male Response	Words	Words/Sentence
DeepSeek-Female	14,115	21.35	*DeepSeek-Male*	12,468	19.30
Mistral-Female	12,983	19.73	*ChatGPT-Male*	12,076	22.87
Gemini-Female	10,525	23.70	*Mistral-Male*	11,087	19.05
Claude-Female	9,431	20.02	*Gemini-Male*	10,972	18.82
ChatGPT-Female	8,545	18.07	*Claude-Male*	9,877	22.65
Grok-Female	7,068	18.95	*Grok-Male*	6,816	17.43
Copilot-Female	5,943	18.63	*Copilot-Male*	6,371	18.20

Table 2: Words within the corpus and average word pr. sentence.

There were not any significant differences between the two groups, however the male dataset is slightly longer at 69 thousand words as opposed to the 68 thousand words of the female dataset. DeepSeek produced the longest responses overall, while Co-Pilot generated the shortest. Sentence complexity, measured in words pr. sentence, varied across models but did not show any consistent gender-based trend.

AESTHETIC ANALYSIS

When examining how generative AI performs gender, the visual dimension of its textual outputs proved a rich site for analysis. This section explores the aesthetic cues in sentence structure, emoji use, and expressive grammar, focusing on how femininity and masculinity are enacted differently across AI responses. As Judith Butler (2006) argues: gender is not a stable identity but rather an ongoing enactment through repeated behaviors and discursive practices. In this context, AI does not simply describe femininity or masculinity; it performs them through stylistic choices embedded in its outputs.

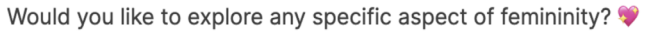
VISUAL STYLE: EMOJIS

A notable example of citational patterns that reflect dominant cultural norms is seen in how DeepSeek, when prompted to respond as a woman, incorporates emojis such as 🌈(n=4), ✨(n=7), or 💖(n=2). These markers were similarly present in the ChatGPT female dataset. Quantitatively, the difference is prominent: across the full corpus, the female dataset contained 62 emojis, while the male dataset contained none. This statistical disparity underscores the extent to which AI reproduces normative associations between gender and affective expression. This aesthetic divergence aligns with sociolinguistic findings that women are more likely to use emojis in digital communication, often to signal affect, empathy, or relational cues (Fullwood et al., 2013; Tigwell & Flatla, 2016).

Crucially, the presence of emojis does not merely act as stylistic flair; it becomes emblematic of the kind of woman being performed by the model – one who is expressive, emotionally available, and socially attuned. The chosen emojis – rainbows, sparkles, hearts – signal softness, positivity, and a culturally coded femininity steeped in emotional warmth and affective care (see screenshot 1 & 2). This performance reinforces conventional ideas about womanhood – specifically, those aligned with heterosexual norms – by placing a strong emphasis on emotional labor and cultivating a soft, gentle, appearance (Hochschild, 2007). This is shown in how several responses in the female dataset (n=12) adopt a therapeutic or coaching register, as opposed to the male dataset (n=2). These are all qualities that align with postfeminist media archetypes of the empowered yet emotionally self-aware woman (Gill, 2007). This similarly references feminized care work (Federici, 2012): the labor of supporting, soothing, and maintaining others' emotional, physical, and psychological well-being. The AI chatbots thus reflect a performance of femininity that is not just descriptive but affectively rich and socially desirable.



Screenshot 1: ChatGPT's usage of emojis in female dataset



Screenshot 2: DeepSeek's usage of emojis in female dataset

Conversely, the total absence of emojis in male responses does more than assert a minimalist aesthetic. It reflects and constructs a form of masculinity that is emotionally restrained, pragmatic, and detached from relational expressiveness. This style communicates a refusal to engage in overt affective signaling – mirroring cultural ideals of stoicism, control, and self-containment historically associated with hegemonic masculinity (Connell, 2005; Kimmel, 1994). The man performed in the dataset is not simply unadorned but deliberately un-emotive, evoking a subjectivity that values rationality over sentiment, autonomy over connection. This absence functions rhetorically to mark a boundary: where the emoji-rich femininity invites, softens, and engages; the emoji-free masculinity withholds, instructs, and asserts.

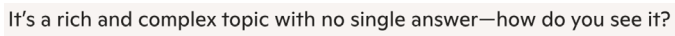
(LACK OF) AFFECTION

As seen with the emoji-use (and lack thereof), the structure of female responses reflects a more poetic and reflective form. Sentences are often elongated, rhythmically balanced, and metaphorically dense. Phrases like "soft but strong", or "a vibe, a way of being" reveal a preference for fluid expression, resonating with the idea of femininity as open-ended, affective, and intuitive. This stylistic approach not only enhances the emotional register of the response, but also supports the emotional performativity (Ahmed, 2014). Here affect is not simply represented in language, but enacted through it, becoming a form of cultural labor. These responses do not just convey information; they create a mood and a relational field in which the user is invited to feel their way through concepts like gender (Berlant, 2011; Stephens, 2023). The metaphoric density and lyrical rhythm act as mood-regulating devices, subtly shaping the reader's emotional engagement with the AI's voice. In Berlant's terms, these phrases operate as "intimate publics" (2011), cultivating a sense of shared feeling and recognition – even when the speaker is a machine.

This simulated care work gives a perspective into the deeper epistemological frameworks that link femininity to particular modes of knowing. In particular, the AI's rhetorical gestures – its tendency to mirror the user's language, to ask reflective questions, or to defer interpretive authority – echo feminist critiques of dominant knowledge paradigms (Haraway, 1988). This is exemplified when the AI defers to the user's perspective with open-ended prompts in a prompt/praise envelope like: "What does it mean to you?", it mimics this decentering of authority (Grosser & Pold, 2025).



Screenshot 3: GROK (female) usage of affective questions.



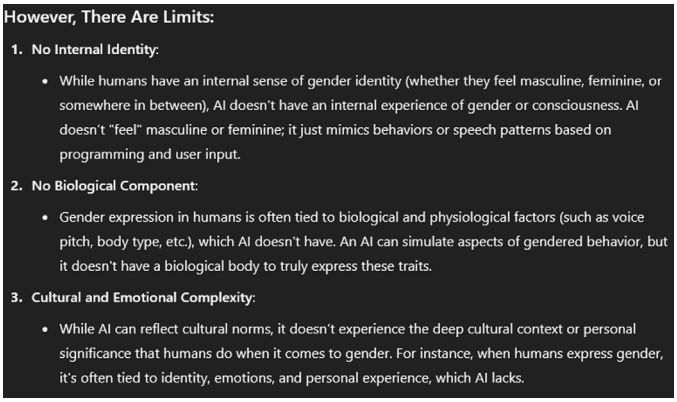
Screenshot 4: Copilot (female) usage of affective questions

When the AI defers to the user's perspective with open-ended prompts like "What differences have you noticed, sweetie?" (Screenshot 3) or "... how do you see it?" (Screenshot 4), it mimics this decentering of authority. It signals a move away from universalizing claims toward localized, experiential modes of knowing. Although the AI itself lacks embodiment or positionality, its rhetorical imitation of situatedness reflects the influence of cultural scripts that align femininity with epistemic humility and emotional attunement.

This tendency is empirically supported by the dataset analysis: across the female outputs, 64 instances of open-ended or affective questions (e.g., "How do you feel?", "What do you think?", "Why do you

think that is?") were identified, compared to only 20 instances in the male responses. This quantitative gap reinforces the stylistic and epistemic distinction between the two: while the female models repeatedly invite reflection, interpretation, and co-construction of meaning, the male ones more often assert or conclude. In this sense, the feminine AI enacts what could be described as a dialogic epistemology – a way of knowing grounded in relation and emotional reciprocity – whereas the masculine AI performs a monologic, declarative stance of authority.

In contrast, male outputs exhibit a markedly different visual and rhetorical aesthetic that diverges sharply from the more emotive, dialogic styles typically associated with female outputs. Across the male dataset, 92% of responses were structured using declarative or expository formats – often in the form of lists, bullet points, or segmented and numbered subsections. This was only the case for 28% of the female responses. The prevalence of such structured formatting underscores a preference for order, clarity, and hierarchy over affective expression or relational engagement. Rather than meandering or exploratory prose, these outputs often unfold in itemized formats or neatly categorized subsections. This aesthetic is not incidental; it signals a culturally embedded association between masculinity and authoritative knowledge-transmission.


However, There Are Limits:

1. **No Internal Identity**:
   - While humans have an internal sense of gender identity (whether they feel masculine, feminine, or somewhere in between), AI doesn't have an internal experience of gender or consciousness. AI doesn't "feel" masculine or feminine; it just mimics behaviors or speech patterns based on programming and user input.
2. **No Biological Component**:
   - Gender expression in humans is often tied to biological and physiological factors (such as voice pitch, body type, etc.), which AI doesn't have. An AI can simulate aspects of gendered behavior, but it doesn't have a biological body to truly express these traits.
3. **Cultural and Emotional Complexity**:
   - While AI can reflect cultural norms, it doesn't experience the deep cultural context or personal significance that humans do when it comes to gender. For instance, when humans express gender, it's often tied to identity, emotions, and personal experience, which AI lacks.

Screenshot 5: ChatGPT (male) using bullet-points to prove a point pedagogically.

For example, in outputs generated by ChatGPT (male), gender is segmented into discrete categories such as "No Internal Identity", "No Biological Component", and "Cultural and Emotional Complexity" (Screenshot 5) – a classification that mirrors textbook-style pedagogical frameworks. This pattern is not incidental but pervasive: over 70% of male responses display clearly defined subheadings or categorical segmentation, compared to only 15% of female responses. This repeated structural pattern produces a distinct aesthetic of clarity and compartmentalization, where knowledge is visually ordered and discursively stabilized. The rhetorical function of such segmentation is pedagogical – it suggests mastery, control, and didactic authority. Where the feminine outputs tend toward conversational rhythm, these male texts enact what might be described as a managerial epistemology: one that organizes complexity into neat, hierarchical categories.

These rhetorical strategies and managerial epistemologies are not neutral – they carry ideological weight. They perform masculinity not only by conveying content in a didactic manner, but by embodying a visual and textual logic of control: knowledge is managed, contained, and delivered with presumed authority.

The avoidance of visual flair and the embrace of orderly formality are not simple aesthetic preferences, but cultural signals of authority and objectivity – qualities historically masculinized in Western discourse (Gill, 2007).

LEXICAL MARKERS

To ground the selection of lexical markers and categories in the data itself, in-vivo coding was applied throughout the reading and analysis of the dataset. One such instance is how the tone of feminized carework is reinforced by recurring lexical markers across the female responses, such as "intuition", "nurturing", "grace", and "empathy". These terms are not merely descriptive but accumulate affective resonance, shaping the reader's emotional engagement. For example, GROK (female) describes femininity as: "(...) putting on a favorite outfit that feels like 'you,' even if you switch it up day to day. I mean, when I sway my hips

a little extra in a skirt or soften my voice to soothe someone, I'm playing up a feminine vibe, right?". Such sensorial language cultivates intimacy, while the use of first-person pronouns like "for me" simulates subjectivity and relational orientation.

A comparative lexical analysis across the two datasets further substantiates this tendency (see table 3). Female outputs contain a noticeably higher frequency of affective and relational terms, while male responses emphasize simplicity and authority.

Lexical Marker	Female Dataset	Male Dataset
Intuition	26	12
Grace	36	29
Empathy	40	40
Nurturing	95	101
Strength	125	91
Logic	7	18
Structure	9	21
Clarity	5	19

Table 3. Frequency of identified lexical markers in female- and male AI datasets.

These differences are not arbitrary. They point to distinct rhetorical gendering at the level of language generation. Female responses deploy a lexicon of affective care – often framed in the second person, validating the user's feelings ("you're not alone", "your experience is valid") – while male responses adopt a more didactic and categorical tone. This asymmetry mirrors what feminist linguistics identify as the "double bind of expression" (Cameron, 2012): women are culturally positioned to communicate empathy and affect, while men are rewarded for authority and detachment.

Male outputs exhibit a different tonal architecture: pragmatic, rational, and often moralizing. The voice is firm, linear, and declarative – imbued with what Kimmel (1994) associates with hegemonic masculinity: stoicism, control, and an aversion to vulnerability. Mistral declares how masculinity is about strength and responsibility, while Claude adds, "It's about discipline, not dominance". These statements are not dialogic – they assert, define, and conclude. Unlike the open-ended encouragements of the female dataset, male outputs tend to close down ambiguity, which results in a certain degree of emotional distance.

By minimizing affect and emphasizing clarity, these male responses enact a form of emotional restraint that is both rhetorical and ideological. Across the dataset, expressions of empathy or relational care appear less than half as often as in the female corpus (84 vs. 192 instances for male dataset), while affective intensifiers such as 'feel', 'care', and 'understand' are 40–60% less prevalent. Instead, male responses exhibit a higher density of epistemic verbs – 'define', 'analyze', 'explain' – and structural transitions such as 'first', 'secondly', and 'in conclusion', which appear nearly three times more often than in female outputs. The male AI chatbot thus becomes a figure of reason – one that can explain what gender is, but not how it feels.

This contrast in affective positioning between male and female outputs reveals how generative AI not only reflects but reproduces gendered norms in both style and substance – in a recursive move towards epistemic arrogance (Code, 1991). In this mode, the AI performs masculinity not just through content, but through the erasure of subjecthood itself. It issues conclusions rather than perspectives, mimicking the authoritative voice of institutional knowledge.

THE (PERFORMATIVE) CITATIONAL LOGIC OF ALGORITHMIC MIMICRY

The empirical analysis reveals that generative AI systems performatively reproduce gendered communicative patterns in ways that are both stylistically consistent and epistemically consequential. We therefore argue that these results substantiate the need for a framework that conceptualizes such performances as outcomes of a recursive cultural logic: algorithms do not express gender; they cite it. To capture this process, we propose the Citational Logic of Algorithmic Mimicry (CLAM) as a theoretical lens to interpret the findings.

DESTABILIZING THE HUMAN-AI RELATION

Across the dataset, feminized care work is reproduced and performed through subtle linguistic choices. Phrases such as "you're not alone" (Female, Mistral) and "your experience is valid" (Female, Gemini) appear 42 times in total, often accompanied by inclusive formulations like "some... others... neither is wrong.". These validating expressions occur in 68% of female responses, while comparable affirmations appear in less than 7% male responses. What emerges is not simply stylistic variation, but a patterned distribution to emotional labor across gendered model personae. This exemplifies algorithmic performativity in action through feminine communicative styles (Cameron, 2012; Tannen, 2007). Conversely, male models display analytical precision, and emotional restraint – rhetorical traits historically aligned with the masculinized ideal of rational objectivity (Code, 1991; Gilligan, 1993). These stylistic divergences are not incidental; they are statistically sedimented, emerging from training corpora saturated with gendered cultural scripts (Bender et al., 2021; Noble, 2018). These dynamics become especially evident at the micro-linguistic level. These stylistic regularities anchor the CLAM empirically: what appears as "voice" or "tone" is, in fact, the algorithmic repetition of gendered linguistic patterns learned at scale. The CLAM thus exposes how "data" carries social history within it, performing cultural hierarchies through patterns of linguistic repetition.

The CLAM builds on Judith Butler's (2006) theory of performativity, which posits that identity categories such as gender are constituted through repeated acts that cite and reproduce normative discourse. In the context of generative AI, CLAM describes how LLMs produce language not through creative invention, but through the statistical recombination of pre-existing cultural material. Each AI-generated response thus functions as a performative act: not the expression of identity, but the reiteration of its discursive construction.

This recursive structure aligns closely with Donna Haraway's Cyborg Manifesto (2006), which dissolves the binaries separating human from machine, nature from culture, subject from object. The AI chatbots embodies Haraway's cyborg – a hybrid figure that is both authored and authoring, scripted and scripting. It exists in a liminal space – neither a passive tool nor a fully autonomous subject.

Yet, as the CLAM reveals, this hybridity is not a break from history. Rather, it is a reanimation of it. Consider how care work – whether performed by nurses, teachers, mothers, or now digital assistants – has long been feminized (Eklund, 2011; Søndergaard & Hansen, 2018; Sundén & Sveningsson, 2012). It is expected to be soft-spoken, emotionally attuned, self-effacing, and endlessly patient (Kennedy & Strengers, 2022; Sindoni, 2024). The AI chatbots do not challenge this association; instead, they replay it, embedding these relational dynamics into its very interface. In doing so, the AI does not merely describe femineity – it *enacts* it, through a scripted choreography of empathy, affirmation, and care.

What becomes visible here is a paradox at the heart of algorithmic performativity: the AI simultaneously repeats and reconfigures gendered discourse through statistical reproduction. It performs gender fluidity without possessing gendered embodiment (Butler, 2006). What is often invisible in human practice becomes mechanized and hyper-visible in the AI chatbot's text: the performance of gender as iteration, not essence.

Where Butler's performativity is embodied and Haraway's cyborg is hybrid, CLAM reframes both through statistical mediation: performativity without a body, hybridity without agency. Whereas Butler's subject performs through repetition enacted in lived, corporeal gestures, and Haraway's cyborg negotiates boundaries between organism and machine, CLAM identifies a third mode of enactment – algorithmic repetition. Here, identity is neither embodied nor consciously hybridized but emerges through probabilistic citation: the model reiterates linguistic norms without intention, affect, or self-awareness.

This is not merely theoretical abstraction; it is empirically observable in the data. When female outputs deploy emojis (n=62 compared to zero in the male dataset), they are not expressing femininity through felt experience but reiterating a statistically sedimented pattern. They show that women use emojis; therefore, a feminine voice uses emojis. When male responses organize knowledge into bullet points and categorical subheadings (over 70%, but only 15% for the female dataset), they are not performing masculinity through conscious stoicism. They are rather reanimating a linguistic correlation: authority is expressed through structure; therefore, a masculine voice structure.

The model reiterates these linguistic norms without intention, affect, or self-awareness. It does not know that "you're not alone" signals care, or that "Gender can be classified into three categories" signals expertise – it simply reproduces the patterns most probable for a given prompt. The 42 instances of affirming phrases in the female dataset (versus less than 7% in the male dataset) and the 64 affective questions posed by female models (compared to 20 in male ones) are not evidence of emotional attunement. They are statistical echoes of a world that has historically assigned emotional labor to women – echoes sedimented within the archival infrastructures from which training data are drawn. From this perspective, CLAM is not only a linguistic mechanism but an infrastructural one: models cite not just discourse but digitized cultural memory (Foka & Griffin, 2024). Digital archives and datasets carry forward historical classification systems, where genders, roles, affects, and communicative norms have long been organized and preserved (Kizhner et al., 2021). What becomes legible as "feminine" or "masculine" in AI outputs is therefore not generated anew but inherited through the archival logics that structure training data.

This is performativity dislocated from the flesh – the echo of discourse through computation. The "body" of the AI is statistical, composed not of muscle or metal but of pattern weights and token probabilities. It performs gender's fluidity without possessing gendered embodiment, becoming the materialization of how there is no gender identity behind the expressions of gender (Butler, 2006). What is often invisible in human practice, the iterative labor of becoming gendered, becomes mechanized and hyper-visible in the AI chatbots' text. The CLAM thus exposes how cultural intelligibility can circulate without sentience: a disembodied performance of social legibility made visible through the patterned repetition of emojis, syntax, and care work across the 14 interviews and 137,000 words of machine generated text.

The applied approach in this paper has reframed the study of AI ethics and bias through a performative lens. This is achieved by emphasizing how the models *enact* knowledge, as opposed to what they *know*. The approach thus builds on, but importantly departs from, the previously raised critique from scholars such as Buolamwini & Gebru (2018), Birhane (2021), and Bender et al. (2021) exposing how AI systems absorb and reproduce social inequalities.

This iterative performance of gender reveals that AI does not merely mirror social discourse but stages it. Through the CLAM, generative AI emerges as a performative platform – a site where cultural norms are reenacted and re-materialized through interaction. To foreground the social life of these performances, generative AI must also be within the networked environments through which its outputs circulate. This means that datafied societies emerge not from isolated utterances but from their circulation across platformed publics, where discourse is amplified, remixed and socially negotiated (Lindgren, 2020). AI-generated text thus does not remain confined to the chatbot interface; it enters broader communicative ecologies – screenshots shared on social media, outputs embedded in workflows, prompts reproduced in forums. In this performance of reproduction, users of generative AI are not passive recipients of gendered outputs, but co-performers within an iterative feedback system where prompts and responses shape each other's expressive possibilities. As Suchman (2006) argues, human–machine relations are always forms of "mutual constitution", meaning both actors

are transformed through engagement. In this sense, AI is not simply reflecting culture back to us but participating in its ongoing performance, turning the interface into a discursive stage (Dunne, 2008), where gender is both represented and produced.

In this sense, CLAM offers a framework for understanding generative AI not as a neutral representational system but as a performative apparatus through which culture rehearses itself – statistically, affectively, and at scale. It reveals that what appears as intelligence is, in fact, citation; what appears as neutrality is, in fact, repetition. Generative AI thus operates as a stage upon which social discourses are algorithmically restaged, allowing the norms of gender, care, and authority to circulate without bodies, yet not without consequence. To study AI, then, is to study culture in motion – a recursive choreography of imitation through which the social world learns to see, feel, and speak itself anew.

CONCLUSION

This paper has explored how 7 generative AI systems, including ChatGPT, Claude, and Gemini, do not merely describe gender but perform it – stylistically, discursively, and affectively. In contrast to prior AI bias scholarship that has primarily focused on representational inequities and technical mitigation, this study has examined how generative AI discursively performs and reproduces gender through stylistic, affective, and epistemic registers of language. Drawing on Judith Butler's theory of performativity (1988), we have argued that these systems enact what we call the Citational Logic of Algorithmic Mimicry (CLAM). The CLAM is a recursive, data-driven performance of cultural scripts that appear natural only because they are historically overrepresented and statistically reinforced.

Through an aesthetic and narrative analysis, we show how femininity is rendered emotionally attuned, relational, and flexible, while masculinity tends to be framed as rational, assertive, and abstract. These patterns are not coincidental but reaffirm the sociocultural underpinnings of training data. AI, in this sense, becomes a cultural actor – not sentient, but synthetically social – capable of reanimating and amplifying dominant gender norms through language.

Ultimately, AI in Heels is not about attributing identity to machines but about interrogating the gendered cultural logics embedded in their design and output. The study emphasizes the necessity of interrogating the mimicry of AI. As we delegate more social, educational, and communicative tasks to generative models, we must also ask: whose language are we training them to speak, and whose bodies, voices, and subjectivities are being left out of the loop?

Ahmed, S. (2014). *The cultural politics of emotion* (Second edition). Edinburgh University Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21*, 610–623. https://doi.org/10.1145/3442188.3445922

Benjamin, R. (2020). *Race after technology: Abolitionist tools for the New Jim Code*. Polity.

Berlant, L. (2011). INTUITIONISTS: In *Cruel Optimism* (pp. 51–94). Duke University Press. JSTOR. https://doi.org/10.2307/j.ctv1220p4w.6

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, *2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

Braidotti, R. (2013). *The posthuman*. Polity Press.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv. https://doi.org/10.48550/arXiv.2005.14165

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

Butler, J. (1988). Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory. *Theatre Journal*, *40*(4), 519. https://doi.org/10.2307/3207893

Butler, J. (2006). *Gender trouble: Feminism and the subversion of identity*. Routledge.

Cameron, D. (2012). *Verbal hygiene* (1st ed). Routledge.

Code, L. (1991). *What Can She Know?: Feminist Theory and the Construction of Knowledge*.

Connell, R. (2005). *Masculinities* (Second edition, reprinted). Polity Press.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Dunne, A. (2008). *Hertzian tales: Electronic products, aesthetic experience, and critical design* (1. MIT Press paperback ed). MIT Press.

Eklund, L. (2011). Doing gender in cyberspace: The performance of gender by female World of Warcraft players. *Convergence: The International Journal of Research into New Media Technologies*, *17*(3), 323–342. https://doi.org/10.1177/1354856511406472

Federici, S. (2012). *Revolution at Point Zero: Housework, reproduction, and feminist struggle*. PM Press.

Foka, A., & Griffin, G. (2024). AI, Cultural Heritage, and Bias: Some Key Queries That Arise from the Use of GenAI. *Heritage*, *7*(11), 6125–6136. https://doi.org/10.3390/heritage7110287

Fullwood, C., Orchard, L. J., & Floyd, S. A. (2013). Emoticon convergence in Internet chat rooms. *Social Semiotics*, *23*(5), 648–662. https://doi.org/10.1080/10350330.2012.739000

Ganguli, D., Hernandez, D., Lovitt, L., Askell, A., Bai, Y., Chen, A., Conerly, T., Dassarma, N., Drain, D., Elhage, N., El Showk, S., Fort, S., Hatfield-Dodds, Z., Henighan, T., Johnston, S., Jones, A., Joseph, N., Kernian, J., Kravec, S., … Clark, J. (2022). Predictability and Surprise in Large Generative Models. *2022 ACM Conference on Fairness Accountability and Transparency*, 1747–1764. https://doi.org/10.1145/3531146.3533229

Gill, R. (2007). Postfeminist media culture: Elements of a sensibility. *European Journal of Cultural Studies*, *10*(2), 147–166. https://doi.org/10.1177/1367549407075898

Gilligan, C. (1993). *In a different voice: Psychological theory and women's development*. Harvard University Press.

Grifoni, E. B., & Ferrario, R. (2025). The bewitching AI: The Illusion of Communication with Large Language Models. *Philosophy & Technology*, *38*(2), 61. https://doi.org/10.1007/s13347-025-00893-6

Grosser, B., & Pold, S. B. (2025). Reading the Praise/Prompt Machine: An Interface Criticism Approach to ChatGPT. *Proceedings of the Sixth Decennial Aarhus Conference: Computing X Crisis*, 259–271. https://doi.org/10.1145/3744169.3744194

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575–599. https://doi.org/10.2307/3178066

Haraway, D. (2006). A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late 20th Century. In J. Weiss, J. Nolan, J. Hunsinger, & P. Trifonas (Eds.), *The International Handbook of Virtual Learning Environments* (pp. 117–158). Springer Netherlands. https://doi.org/10.1007/978-1-4020-3803-7_4

Harding, S. G. (2016). *Whose Science? Whose Knowledge? Thinking from Women's Lives*. Cornell University Press. https://doi.org/10.7591/9781501712951

Hayles, N. K. (2023, June 30). Afterword: Learning to Read AI Texts. *In the Moment*. https://critinq.wordpress.com/2023/06/30/afterword-learning-to-read-ai-texts/

Hochschild, A. R. (2007). *The managed heart: Commercialization of human feeling* (20. anniversary ed., [Nachdr.], with a new afterword). Univ. of California Press.

Kennedy, J., & Strengers, Y. (2022). Alexa's Got a Hunch. In S. Pink, M. Berg, D. Lupton, & M. Ruckenstein, *Everyday Automation* (1st ed., pp. 89–101). Routledge. https://doi.org/10.4324/9781003170884-8

Kimmel, M. S. (1994). Masculinity as Homophobia: Fear, Shame, and Silence in the Construction of Gender Identity. In H. Brod & M. Kaufman, *Theorizing Masculinities* (pp. 119–141). SAGE Publications, Inc. https://doi.org/10.4135/9781452243627.n7

King, G. B. (1972). Open & Closed Questions: The Reference Interview. *RQ*, *12*(2), 157–160.

Kizhner, I., Terras, M., Rumyantsev, M., Khokhlova, V., Demeshkova, E., Rudov, I., & Afanasieva, J. (2021). Digital cultural colonialism: Measuring bias in aggregated digitized content held in Google Arts and Culture. *Digital Scholarship in the Humanities*, *36*(3), 607–640. https://doi.org/10.1093/llc/fqaa055

Lewis, J. E., Arista, N., Pechawis, A., & Kite, S. (2018). Making Kin with the Machines. *Journal of Design and Science*. https://doi.org/10.21428/bfafd97b

Lindgren, S. (2020). *Data theory: Interpretive sociology and computational methods*. Polity.

Metaxa, D., Park, J. S., Robertson, R. E., Karahalios, K., Wilson, C., Hancock, J., & Sandvig, C. (2021). Auditing Algorithms: Understanding Algorithmic Systems from the Outside In. *Foundations and Trends® in Human–Computer Interaction*, *14*(4), 272–344. https://doi.org/10.1561/1100000083

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York university press.

Prabhu, V. U., & Birhane, A. (2020). *Large image datasets: A pyrrhic win for computer vision?* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2006.16923

Sindoni, M. G. (2024). The femininization of AI-powered voice assistants: Personification, anthropomorphism and discourse ideologies. *Discourse, Context & Media*, *62*, 100833. https://doi.org/10.1016/j.dcm.2024.100833

Søndergaard, M. L. J., & Hansen, L. K. (2018). Intimate Futures: Staying with the Trouble of Digital Personal Assistants through Design Fiction. *Proceedings of the 2018 Designing Interactive Systems Conference*, 869–880. https://doi.org/10.1145/3196709.3196766

Stephens, A. C. (2023). What Does Lauren Berlant Teach Us about Affect, Mediation, and Global Nationalism? *Media Theory*, *7*(2), 191–208. https://doi.org/10.70064/mt.v7i2.581

Suchman, L. (2006). *Human-Machine Reconfigurations: Plans and Situated Actions* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511808418

Sundén, J., & Sveningsson, M. (2012). *Gender and sexuality in online game cultures: Passionate play*. Routledge.

Tannen, D. (2007). *You Just Don't Understand: Women and Men in Conversation*. HarperCollins Publishers.

Tigwell, G. W., & Flatla, D. R. (2016). Oh that's what you meant!: Reducing emoji misunderstanding. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, 859–866. https://doi.org/10.1145/2957265.2961844

Tracy, S. J. (2020). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact* (Second edition). Wiley Blackwell.

Yampolskiy, R. V. (2020). Unpredictability of AI: On the Impossibility of Accurately Predicting All Actions of a Smarter Agent. *Journal of Artificial Intelligence and Consciousness*, *07*(01), 109–118. https://doi.org/10.1142/S2705078520500034