

THE ONTOLOGICAL AMBIVALENCE OF ARTIFICIAL INTELLIGENCE IN GENDER STEREO TYPING

By

EMMANUEL INIOBONG ARCHIBONG, Ph.D.

Department of Philosophy

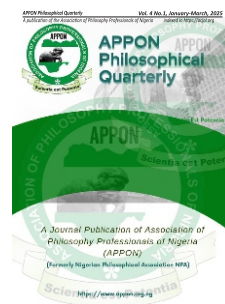
University of Uyo, Uyo.

Mobile: +2348032325087

E-mail: emmaiarchibong@uniuyo.edu.ng



Copyright: © 2025 by the authors. Submitted
for possible open access publication under the
terms and conditions of the Creative Commons
Attribution (CC BY) license
(<https://creativecommons.org/licenses/by/4.0/>)



ABSTRACT

Artificial intelligence (AI) systems are influencing human lives but with greater potential to perpetuate or challenge gender stereotypes as a pressing concern. The ontological ambivalence of AI gender stereotyping refers to the inherent contradictions and tensions within these AI systems, which can both reflect and shape societal norms. Despite the growing recognition of AI's impact on gender equality, the ontological ambivalence of AI gender stereotyping remains understudied. This study aims to critically examine the ontological ambivalence of AI gender stereotyping and its implications for individuals and society. The fundamental research question is: "How do AI systems perpetuate gender stereotypes, and what are the implications for individuals and society?" To achieve this aim, the study analyse existing research on AI, gender stereotyping, and ontological ambivalence and examines AI-generated content and algorithms to identify patterns and contradictions in gender stereotyping. The findings suggest that AI systems exhibit ontological ambivalence, reflecting both traditional and progressive values, and that these systems can perpetuate and challenge gender stereotypes in complex ways. It concluded that understanding the ontological ambivalence of AI gender stereotyping is important for developing effective strategies for challenging and subverting limiting gender norms by developers of AI systems. The study recommended that developers prioritize fairness, accountability, and transparency in AI systems, and that policymakers implement regulations to prevent AI-generated gender stereotypes from perpetuating harm.

Keywords: Artificial Intelligence, Gender Stereotyping, Ontological Ambivalence, Societal Norms, Fairness, Accountability.

Introduction

Artificial Intelligence (AI) systems are built using algorithmic logic, probabilistic modelling, and pattern recognition. These are not neutral technical processes as they constitute the ontological superstructure of AI which is a worldview embedded in code, dictating what can be known and how. Gender is a deeply contextual, socially

performed identity and clashes with this structure. Instead of fluid representation, AI seeks computable form; instead of social distinction, it demands categorical input. This misalignment lies at the heart of what this study terms the *ontological ambivalence* of AI gender stereotyping harping on the structural tension in which AI systems simultaneously reinforce and

obscure gender norms.

Rather than reflecting reality, AI constructs it through classification schemas that encode ontological commitments. When gender is forced into binary variables for model training, the result is not a passive misrepresentation but an active construction of gender. Judith Butler has argued that gender is not a stable identity but a stylised repetition of acts, a performativity continually realised through discourse and power structures.¹ AI, however, is built on statistical aggregation and binary classification, both of which negate the performative instability Butler describes. Its ontology demands a fixed corpus, thereby annihilating gender into a predictable, processable form. The implication is that, gender becomes an artefact of algorithmic sorting. For instance, a machine learning model trained on hiring data will not merely pick up male-dominated job titles as it will construct "male" as the statistical norm for certain professional identities. The system does not copy bias; it learns an ontology where gendered roles are causally real. The consequence is that women, trans, and non-binary individuals are often misread, misrepresented, or excluded not because the algorithm fails, but because it functions precisely as it was designed: to generalise. Thus, gender stereotyping is not an unfortunate accident of AI but a *structural feature* of its being. Ambivalence enters at the point where AI seems capable of both progress and regression. For instance, the case of AI-generated image platforms that now include gender-neutral options or predictive text that corrects gendered language. These features appear inclusive but underneath them often lies a statistical model trained on biased corpora where

normative gender roles remain dominant. The inclusivity is therefore superficial and is added on top of a system whose ontology remains binary and essentialist. The contradiction between these layers, an inclusive surface and a reductive core is the ontological ambivalence this study investigates. This ambivalence is more than ethical uncertainty; it is a structural contradiction. Algorithms are built to optimise, and optimisation prefers regularity, patterns, and categories. Gender, however, thrives on context, variability, and social negotiation. The AI system must thus either force gender into simplified schemas or exclude it entirely both of which result in ontological harm. What emerges is a system that protects certain expressions of gender while marginalising or erasing others, depending on how legible they are to the algorithm's internal logic.

In attempting to protect female users from harassment, many platforms employ automated filters that detect gendered slurs and targeted abuse. However, these systems frequently flag LGBTQ+ content or feminist speech as abusive, particularly when it includes reclaimed or colloquial expressions. The algorithm protects gendered identity only when that identity aligns with its trained logic.² Non-normative or intersectional expressions are often unrecognised or over-policed. What appears to be a form of ethical sensitivity is, in fact, a function of ontological rigidity. Miranda Fricker has described the condition in which individuals are misrepresented or silenced within knowledge systems as epistemic injustice, particularly testimonial and hermeneutical forms.³ In algorithmic environments, this becomes ontological injustice: systems that fail to register non-

¹ Judith Butler, *Gender Trouble: Feminism and the Subversion of Identity* (New York: Routledge, 1990), 25.

² Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press, 2018), 139.

³ Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press, 2007), 1.

binary or fluid gender identities as real, thereby excluding them from the computational ontology altogether. Here, the injustice is not only about representation but about existence within the system what AI “knows” and what it can “see” becomes a gatekeeping mechanism for being recognised as real. Joy Buolamwini and Timnit Gebru analysed commercial facial recognition tools and discovered stark disparities in classification accuracy: the error rate for identifying dark-skinned women was over 34%, compared to less than 1% for light-skinned men.⁴ The technical explanation points to unbalanced training datasets, but the philosophical implication is deeper: the system's ontology of the human is racially and gendered incomplete.

What Is Ontology in AI?

Ontology, in its classical and contemporary sense, interrogates not just what exists but how things exist and how they are categorized within a system of knowledge. In artificial intelligence, this metaphysical foundation becomes materialised in algorithmic forms which are rules, labels, training data, and model architectures. Unlike the open-ended and context-sensitive structures of natural human cognition, AI systems depend on rigid ontologies to function. They do not merely process representations of the world, they instantiate a worldview. Every act of classification and recognition by an AI system implies an ontological choice such as, this is this kind of thing, and not another. In gendered terms, this comes out as a binary logic of recognition that fails to grasp the fluid, lived, and contested nature of human gender experience.

The philosophical character of AI thus

becomes inseparable from its technical design. Emmanuel Archibong points out that ontology “aims to provide answers to issues concerning the nature of entities and the ways in which they can be classed, connected, and organized under a reality framework”. With regards to AI, these classifications are computationally implemented such as entities are rendered into datasets and ontologies that the system can manipulate. The danger lies in the fact that such formalisation requires simplification. Gender, instead of being a process of becoming or a site of performance, is rendered into a static variable often binary, sometimes a fixed list, always simplified for operational clarity. These reductions are not neutral as they encode assumptions about the nature of gender that are embedded in the system's architecture.

The problem becomes evident in practices such as natural language processing and computer vision, where gender is either explicitly tagged or implicitly inferred. A chatbot that refers to a user as “he” or “she” based on name recognition is not only making a pragmatic inference but is declaring an ontological truth about the user's identity. That inference may be statistically sound but philosophically flawed. The classification is premised on a computational ontology that cannot accommodate ambiguity, change, or resistance to categorisation. What results is not just a mistaken output but also a reduction of being: which the system denies ontological space to those who live outside its conceptual network.

The stakes of this reduction are particularly clear when one examines the logic of supervised machine learning. In

⁴ Joy Buolamwini and Timnit Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification,” *Proceedings of Machine Learning Research* 81 (2018): 5.

⁵ Emmanuel I. Archibong, *The Ontology of the Mind's Creativity and the Language of Artificial Intelligence*, USEM Journal of Languages, Linguistics & Literature, Vol. 12 (2024): 165.

order to train a model, data must be annotated of which examples must be labelled according to categories predefined by human annotators. In gendered contexts, this might involve tagging faces or names as “male” or “female.” These tags become the truth-values the model learns from. Over time, the system internalises a reality in which these gender categories are not just statistically frequent, but ontologically primary. The AI is not only trained to detect differences but to believe in their fixed status. Such a design produces a system in which the very ability to see or acknowledge nonbinary or fluid gender identities is absent. It is not that the AI is biased in the human sense but that its ontology does not allow for the existence of those it cannot categorise.

This is why the distinction between ontology and ethics is critical. Much discussion on AI bias focuses on ethical dimensions such as fairness, representation, and harm reduction. But fairness audits do not change the system’s ontological assumptions. One can reduce disparate outcomes without ever questioning the categories through which those outcomes are derived. Ontology asks for instance, “what kinds of beings does the system know?” What does it mean for something to be real in the system’s world? These are not questions of error but of design. They are not concerned about how well the system performs, but with what kind of reality it constructs through its functioning.

The AI system, by necessity, enforces closure. It cannot function without a completed ontology as each entity it processes must belong to some class, and each input must resolve into a usable category. But human gender is inherently resistant to closure. It is defined by performative acts, social interpretations, internal subjectivity, and external

negotiations. The insistence of AI on resolving gender into a stable form is therefore not just technically expedient but is ontologically violent as it refuses to let gender be what it is, a dynamic field of becoming.

John Searle’s theory of intentionality famously linked mental states to directedness, that is, the mind’s capacity to aim itself toward objects, ideas, and representations.⁶ AI systems, lacking consciousness, possess no such intentionality. They do not mean anything; they process inputs and generate outputs. Yet their lack of intention does not make them inert. On the contrary, they act with profound ontological force. Their outputs are often perceived by users as meaningful, even authoritative. When an AI labels a person or interprets their gendered behaviour, it produces knowledge that others consume as truth. In this way, the system’s lack of subjective experience paradoxically amplifies its ontological power: it does not know that it misrepresents, and therefore cannot be made to care.

This absence of intentionality is not a safeguard though as it is a limitation that supports the ontological closure AI demands. AI cannot ask whether its categories are valid and cannot doubt, revise, or reflect. These are uniquely human capacities tied to subjective self-awareness. As a result, AI’s categories become self-perpetuating. They shape further data collection, influence human behaviour, and surround themselves in institutional practices. For example, if an AI system classifies images of women primarily in domestic contexts, that classification becomes training data for future systems. The system does not merely mislabel but reshapes the world to match its categories.

It is important to understand that AI’s

⁶ John Searle, “Minds, Brains, and Programs,” *Behavioural and Brain Sciences* 3, no. 3 (1980): 419.

ontological rigidity is not accidental but demanded by its technical structure. Algorithms require consistency, and consistency demands simplification. Gruber famously defines ontology in computer science as “an explicit specification of a conceptualization”.⁷ This makes this logic clear which is that the world must be rendered into a formal structure in order for AI to act upon it. But the act of rendering is also an act of reducing. Gender, once modelled, becomes a static object, a computational entity that can be manipulated, predicted, and controlled. In philosophical terms, this represents a shift from being to objecthood, a process of de-ontologising the self into a feature path.

Thus, the way AI systems are programmed at the level of their ontology ensures that gender will always appear in stereotypical or overly simplistic forms. Not because designers are necessarily prejudiced, but because the systems themselves cannot handle the ontological indeterminacy that gender entails. The only genders that will be protected, recognised, or legitimised are those that fit the system’s preordained structure. Others will be excluded not by oversight, but by ontological impossibility.

The Paradox of AI’s Ambivalence

Ambivalence, by its very structure is the state of holding contradictory positions simultaneously such as an affirmation and negation, recognition and denial, protection and erasure. Within the context of AI, ontological ambivalence emerges not from indecisiveness, but from the structural contradiction between AI’s classificatory logic and the reality of human gender complexity. Gender cannot be accurately computed without being categorised, and it cannot be categorised without being simplified. This contradiction is not resolved

in practice but embedded into the very fabric of AI systems.

AI systems operate under rules of decision which is this or that, male or female, acceptable or not. These are not merely technical dichotomies but ontological choices that define what kinds of beings are legible to the machine. Yet at the same time, the system’s output often signals a gesture toward inclusivity such as gender-neutral pronouns, diversity filters and fairness metrics. Thus, AI appears to both respect and restrict gender variation. This duality forms a paradox, one in which the surface signs of fairness mask an underlying logic of exclusion. In this way, ambivalence is not a weakness of the system, but its central mode of functioning.

Voice assistants often default to feminised voices in subservient roles. While this reflects existing social stereotypes, the system itself has no awareness of gender norms. It simply replicates what has proven effective or familiar in user interactions. The feminised assistant is both a product of user preference and of historical data which both reinforce and is shaped by cultural expectations. AI does not intend to stereotype, yet it must rely on stereotypical data in order to function accurately. What seems like ambivalence that is, both promoting and resisting bias is in fact the structural result of ontological design.

This ambivalence is further complicated when AI systems are deployed with fairness considerations. A hiring algorithm might be adjusted to reduce the disadvantage faced by women, for instance, by recalibrating scoring functions. But the very act of “protecting” a gender category presumes its stability. Protection becomes possible only when gender is seen as fixed, detectable, and classifiable. This way, efforts to correct bias often support the very essentialism they seek

⁷ Thomas R. Gruber, “A Translation Approach to Portable Ontology Specifications,” *Knowledge Acquisition* 5, no. 2 (1993): 199.

to escape. The AI recognise recognises only the corrective coefficient added to a binary structure. What results is not liberty, but the reinforcement of categorisation under a benevolent guise.

Uduk and Archibong, in their philosophical treatment of gender stereotyping through the lens of Buber's complementarity, argue that gender should be understood not in terms of superiority or fixed roles, but as a dialogical interaction "a two-way relationship where one, the 'I', encounters the other, 'Thou', with mutual awareness".⁸ The AI system, by contrast, does not enter into dialogue but into classification. It sees the user as an object, not as a Thou. Ambivalence arises here from the ethical pressure to design inclusive systems. This sorting mechanism does not leave room for ambiguity. AI cannot process hesitation, contradiction, or unfolding identity. A system trained to infer gender based on images or text, for example, must make a decision even when uncertainty would be more respectful. There is no "maybe" in machine logic as the result is a forced ontological closure which the system must decide, even when no decision should be made. This imposition of legibility where none exists mirrors what philosopher Miranda Fricker describes as a form of hermeneutical injustice, where individuals are harmed because their social experiences are not adequately understood or represented within dominant interpretive frameworks.⁹

Even when AI systems are designed with explicit gender neutrality in mind, they still rely on inherited ontological categories. The attempt to build "gender-blind" algorithms often results in systems that simply ignore gender complexity altogether, which

silences rather than liberates. The ambivalence here is not between bias and fairness, but between recognition and removal. The system neither enforces stereotype nor affirms complexity but simply deletes gender from its ontological field. This is ontologically reducing the richness of human identity to the limits of what the system can ignore.

The paradox is further amplified in legal and regulatory settings. AI systems tasked with enforcing equality laws or monitoring hate speech often rely on classifiers that must recognise gendered language. In order to protect, they must first identify; in order to identify, they must categorise; and in order to categorise, they must fix what should remain open. Thus, the legal intention of inclusion becomes an ontological imposition. The system must act as if gender is knowable when, in many cases, it is precisely the unknowability of gender that needs to be respected.

Attempts to resolve this ambivalence through technical design have been largely superficial as fairness metrics, gender-balanced datasets, and inclusive interface features offer surface-level responses to what is fundamentally an ontological dilemma. These measures do not change the fact that AI must categorise in order to operate. And every act of categorisation is a narrowing of being. Therefore, even the most well-intentioned AI design reproduces ambivalence, not because its creators are indifferent, but because its structure demands contradiction.

Gender and Stereotyping

Gender stereotype is not merely a biased statement or belief but a socially institutionalised expectation that links

⁸ Happiness E. Uduk and Emmanuel I. Archibong, *Buber's Complementarity and Gender Stereotyping* (Port Harcourt: Gift Austine Company, 2024), vii.

⁹ Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press, 2007), 6.

certain attributes, behaviours, or roles to specific gender identities. These expectations are not just cultural relics but are mechanisms of control and regulation. They tell people who they are allowed to be and how they are expected to perform that being. AI systems that learn from historical and real-time data do not encounter gender in a vacuum. Instead, they are exposed to these performative codes, often without the tools to question or contextualise them. In absorbing patterns, they absorb stereotypes; in replicating outputs, they replicate normativity.

Stereotyping operates ontologically as a mode of fixing identity, that is, making fluid identities static and recognizable. In both human perception and algorithmic computation, stereotypes function as cognitive shortcuts. For human cognition, these shortcuts may be unconscious and emotionally conditioned. For AI, however, the shortcut becomes structural as the machine is not simply choosing the easiest path but executing logic that is hard-coded into its model by associating X term or image with Y label, based on probability derived from training data. When that data reflects gendered assumptions “nurses” as women, “engineers” as men, the model learns to support them. In turn, the stereotype is no longer just a cultural assumption but has become a statistical truth within the system’s decision-making apparatus

Once a stereotype is embedded into the system, it becomes increasingly difficult to dislodge because the model continually strengthens it through feedback loops. This is especially true in systems that interact with large volumes of user-generated content. Recommendation engines, for example, might begin to gather user

preferences in gendered ways, not because users are inherently gender-biased, but because the system learns patterns that presume gender identity based on surface-level behaviour. Stereotyping is not only about visibility but about who is allowed to exist intelligibly within the system’s framework. If the training data lacks representation of trans, non-binary, or gender fluid individuals, or worse, mislabels them, then the system will reproduce this absence in its operational ontology. These users become computationally illegible, and illegibility becomes a form of ontological violence. AI systems then function not only as mirrors of social bias but as engines of normative closure.

In their philosophical account of gender categories, Uduk and Archibong argue that the binary model of male and female has become “a battlefield of superiority complexes and stereotype formations,” resulting in structural antagonism instead of complementary coexistence.¹⁰ The AI system, trained on such binary assumptions, amplifies these divisions by hardcoding them into its representational schema. Where a human might learn, question, and revise assumptions, an AI cannot reflexively interrogate the limits of its gender logic. One telling example is in language generation models. These systems are trained on corpora that reflect years, if not centuries, of gendered discourse. When prompted to write about leadership or intelligence, these models often skew male; when prompted about care, beauty, or nurturing, they belittle female. While such systems can be fine-tuned to avoid obvious offensive content, their internalised ontological model of gender remains largely unchanged.

Such ontological reduction results in pre-emptive epistemology. The AI has

¹⁰ Happiness E. Uduk and Emmanuel I. Archibong, *Buber’s Complementarity and Gender Stereotyping* (Port Harcourt: Gift Austine Company, 2024), vii.

already decided what gender means before it interacts with the user. In this way, the stereotype is not inferred but imposed. The system's outputs are not neutral predictions; they are expressions of learned essentialisms. Even when AI systems attempt to accommodate gender diversity, their strategies often fail to escape stereotyping. Gender options in surveys or platforms, for instance, may include "Other" or allow custom input. However, in the backend, these responses are often mapped back onto a binary for the sake of model simplicity.

One might object that such limitations are not intentional but structural. And this is precisely the point. The AI does not stereotype because it chooses to, but because it must classify. In classification, simplification is inevitable. Stereotyping is not an accident but a function. In this sense, the ontological structure of AI reifies stereotypes by demanding ontological closure at the point where social reality demands openness. Judith Butler warned that treating gender as essence rather than act is a mistake with political consequences.¹¹ AI systems make this mistake daily not through malice, but through design. They turn fluid identities into objects of classification, and in doing so, they institutionalise the very stereotypes that critical theory has long sought to dismantle.

Algorithm and Gender Bias

At the core of gender bias in AI lies the algorithm which is a procedural mechanism that processes data, extracts patterns, and applies learned functions to new inputs. While algorithms are often presented as abstract mathematical tools, they are deeply situated in cultural, political, and

ontological assumptions. They do not function in neutral space but translate a worldview into executable logic. This translation process is where bias is embedded not as a flaw to be corrected, but as a function of what the algorithm is trained to recognise, reward, and replicate.

Most machine learning algorithms rely on historical data to build predictive models. If that data reflects gendered disparities such as wage gaps, occupational segregation, or social expectations, then the resulting algorithm will encode and reproduce those patterns. The issue is not that algorithms consciously discriminate, but that they operate through generalisation. Generalisation itself is a political act when applied to social identity. Through probabilities, algorithms naturalise inequality as a pattern of the world, rather than a problem to be contested. Algorithm can infer that maleness was a proxy for competence.¹² The algorithm was not malfunctioning but was functioning exactly as designed, optimising for patterns of past success.

Algorithms do not merely reflect gender bias but infer gendered patterns and then apply them in novel contexts. Once these inferences are embedded, they are difficult to unearth because they often manifest in subtle ways. Scholars like Cathy O'Neil have argued that algorithmic models can function as "weapons of math destruction," particularly when their outcomes are opaque, unregulated, and self-perpetuating.¹³ In her analysis, predictive algorithms often create negative feedback loops. A person denied credit or employment due to algorithmic scoring may find their future opportunities further constrained, strengthening the initial assessment. In

¹¹ Judith Butler, *Gender Trouble: Feminism and the Subversion of Identity* (New York: Routledge, 1990), 32.

¹² Jeffrey Dastin, "Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women," *Reuters*, October 10, 2018.

¹³ Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016), 28.

gendered terms, women who are statistically underrepresented in leadership positions may be scored lower by an algorithm predicting “executive potential,” thereby preventing the very progress needed to correct the imbalance.

Caliskan, Bryson, and Narayanan demonstrated that word embedding, a foundational technique in natural language processing NLP learn human-like biases simply by being exposed to text. For example, embedding associated “man” with words like “computer programmer” and “woman” with “homemaker”.¹⁴ These biases are not added maliciously but are statistically inferred from frequency patterns in training data. However, when used in downstream tasks such as résumé screening or automated writing, they reproduce gendered divisions at scale.

Even when developers implement moderation layers or fine-tuning to mitigate biased outputs, the underlying model retains patterns it absorbed during retraining. These residual biases can surface in subtle ways. A model might consistently associate negative sentiment with female-coded names, or default to male pronouns in contexts of power and authority. The result is an AI that appears polite and neutral but thinks in stereotypes. There is also a dimension of epistemic bias in how data is selected, cleaned, and labelled. The construction of training datasets often reflects existing social hierarchies. For instance, image datasets used to train vision algorithms may contain disproportionately white, male, Western imagery. In contexts where gender is labelled, such labels are usually binary, ignoring trans and nonbinary identities.

Researchers working with such datasets must often make ontological decisions such as what counts as a “woman”?

Even technical interventions to reduce bias often fall short because they fail to address the structural constraints of the algorithmic paradigm. However, they still operate within a framework that treats gender as a measurable, enumerable attribute. In doing so, they preserve the ontological commitment to classification and legibility that produces stereotyping in the first place. Sandra Wachter and Brent Mittelstadt have critiqued this tendency in legal and policy frameworks, noting that most algorithmic fairness measures are built on “paradigms of individual rationality and formal equality,” which are ill-suited to addressing the contextual, structural, and relational nature of gender inequality.¹⁵ What algorithms lack is the capacity for interpretive distinction which is the ability to recognise how a gendered expression might be shaped by culture, power, or intersectionality.

A 2018 study found that error rates for dark-skinned women reached as high as 34%, compared to less than 1% for light-skinned men.¹⁶ These disparities are not just technical limitations; they are the result of training data that over represents dominant groups. The algorithm learns who counts as “human” by being trained on those who are most visible in its datasets. Gender, in this framework, becomes an aesthetic norm akin to something that can be read from a face, measured by geometry, and judged against a canonical form.

Compounding this problem is the

¹⁴ Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan, “Semantics Derived Automatically from Language Corpora Contain Human-Like Biases,” *Science* 356, no. 6334 (2017): 183–186.

¹⁵ Sandra Wachter and Brent Mittelstadt, “A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI,” *Columbia Business Law Review*, no. 2 (2019): 494.

¹⁶ Joy Buolamwini and Timnit Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification,” *Proceedings of Machine Learning Research* 81 (2018): 5.

commercial imperative driving most AI development. Companies building AI tools are profit-driven to optimise for scale and efficiency, not philosophical coherence or social justice. As a result, datasets are scraped for volume rather than curated for representational accuracy. Models are trained to predict outcomes, not question the premises of those outcomes. In the guise of efficiency, complexity is sacrificed and gender is one of the first complexities to be reduced to the rubbles. The challenge, then, is not only technical but philosophical. Can we build systems that acknowledge the indeterminacy of identity? Can algorithms learn to operate without fixed categories?

Ontological Balance of Artificial Intelligence and Gender Equity

AI systems present a paradoxical situation in the politics of gender as they can simultaneously act as instruments of social reproduction and vehicles of disruption. At the centre of this paradox is an ontological tension through AI's need to impose structured knowledge onto unstructured realities. This tension reveals itself most clearly in the competing impulses within AI systems: to recognise and reinforce dominant gender norms, or to disrupt and challenge those norms through alternative representations and relations. However, this is not a matter of surface output alone. The underlying logic of the algorithmic system which is the way it categorises, generalises, and makes legible shapes what gender becomes within AI's operational frame.

What is most striking about this tension is that it does not originate from malice or intention. AI systems do not possess ideological commitments as they inherit them from the data and social systems that surround their development. When trained on historical datasets, AI learns the shape of gender as it has been structured through decades of social relations, media representation, economic participation, and institutional practice. These inherited

patterns are often unequal and binary. Thus, the system defaults to perpetuation not because it is biased in a human sense, but because perpetuation is the most efficient and legible form of prediction.

An AI cannot easily account for identities that shift, blur, or multiply, especially when these identities lack statistical prominence in the training data. The result is a system that is structurally unprepared for ontological pluralism. Even when designers intend for inclusivity, they confront technical limitations that trace back to fundamental assumptions about what counts as knowable. The tension arises because AI models are designed around classification, and classification is inherently reductive. Yet gender is not a reducible phenomenon. It is shaped by cultural location, embodied experience, historical change, and political struggle. A model that classifies "female" based on facial features or linguistic patterns does not reflect gender but constructs a version of gender that conforms to its internal ontology. In this construction, the potential for transformation is subordinated to the demand for stability.

True transformation would require an AI that can function without reducing gender to a fixed attribute. This may involve abandoning or radically altering classification models altogether. It would also mean engaging with interdisciplinary knowledge from gender studies, critical race theory, anthropology, and philosophy to build systems that understand gender as relation, context, and performance, rather than essence. Few AI systems are currently designed this way, in part because doing so introduces complexity that resists automation. But without such ontological humility, AI systems will continue to reproduce existing hierarchies, even when marketed as tools for equity.

Kate Crawford and Trevor Paglen have explored the visual politics of AI training data, showing how massive image datasets

construct “normative categories of personhood” by privileging certain facial structures, poses, and aesthetic cues.¹⁷ Their work highlights how gender, race, and class are not just variables but constructed through the visual logic of the machine. In these systems, being recognised is not a neutral act but an act of power. If recognition is granted only to those who match the model’s learned norm, then those outside it become unnoticed, uncountable, and, effectively, unrecognisable within the system’s ontology. This logic of recognition explains the limits of current AI diversity efforts. Including more data from marginalised groups is important, but insufficient. Representation alone does not undo ontological assumptions. If the system still learns gender as a stable category based on appearance or speech, then it will continue to stereotype, even if it now stereotypes a broader range of people. What is needed is not better representation within a flawed ontology, but an interrogation of the ontology itself.

That interrogation begins by asking whether gender should be treated as a target variable at all. In many contexts, gender inference is unnecessary or harmful but it persists because it serves institutional and commercial purposes. The more AI learns to use gender as a computational resource, the less space remains for gender to be lived outside of norms. To move beyond this impasse, AI development must engage with concepts of ontological multiplicity which is the idea that being is not singular or reducible but multi-layered, situated, and dynamic. This idea, drawn from post-structuralist and decolonial thought, brings about a radical reimagining of what it means to build “intelligent” systems. It calls for systems that can encounter ambiguity without resolving it, that can hold

contradiction without collapsing it into error. Such systems would not only support gender transformation but would be transformed by it. As long as AI continues to treat gender as a static input for predictive modelling, it will remain biased. But when AI is challenged to engage with gender as a site of on-going becoming, it may begin to support the kind of transformation that gender justice requires. That transformation cannot come from code alone but must arise from a reorientation of the very questions we ask when we build and deploy intelligent systems.

Evaluation

From the ensuing discussion, it can be noted that Artificial intelligence has often been heralded as a mechanism for social progress, capable of extending fairness, enhancing objectivity, and eliminating human error. But when it comes to the issue of gender, these claims fall short of expectations. AI as has been discovered is built upon a classificatory logic that does not merely reflect gendered realities but actively reconstructs them according to statistical generalisations. The most immediate and dangerous failure of AI systems in this context is their ontological rigidity which is the idea that identities, including gender, must be fixed, classifiable, and computationally stable.

These tools measure the system’s ability to reproduce expected outcomes but what if the expectations themselves are sexist, transphobic, or otherwise exclusionary? While these tools aim to address inequality, they rarely interrogate the categories on which these metrics depend. They assume gender to be a binary, static attribute and a label affixed to data subjects rather than a lived, constructed, and relational identity. This means the system can be adjusted to

¹⁷ Kate Crawford and Trevor Paglen, “Excavating AI: The Politics of Images in Machine Learning Training Sets,” *Excavating AI*, 2019.

treat men and women “equally” according to the data, while entirely erasing non-binary people or those whose gender identity shifts across contexts. Intersectional feminist scholars propose that treating AI systems as neutral tools to be optimised, and as political artefacts that must be understood in terms of power, history, and positionality. This becomes a method of uncovering whose knowledge counts, whose identities are legitimised, and whose lives are rendered computationally expendable.¹⁸

Conclusion and Recommendations

The ontological ambivalence of AI gender stereotyping lies not in its occasional failures or biased outputs, but in the fundamental structure of how these systems come to “know” and act upon gender. From the arguments in this study, it can be observed that AI does not merely process gender but reconstructs it, often through binary, essentialist, and reductive logics embedded in algorithmic classification. The question, therefore, is not whether AI systems contain bias, but whether their very ontology permits them to engage with gender in ways that are just, plural, and open-ended. The conclusion is stark: under current architectures, they do not.

AI systems function through categorisation, generalisation, and prediction. Gender, however, defies stable classification. It is a lived, social, and relational phenomenon that exceeds any closed computational model. When AI is tasked with interpreting or predicting gender whether through names, faces, speech, or text it is forced to reduce this complexity to pre-defined variables. The system cannot acknowledge the fluidity of gender because its design does not allow for ambiguity. Instead, it stereotypes, simplifies, or erases

all under the guise of efficiency.

In light of these findings, recommendations must go beyond surface-level technical interventions and target the philosophical, architectural, and procedural foundations of AI design. These recommendations are not just for developers but also for regulators, educators, researchers, and users. Automated inference of gender through names, appearance, or voice should be strongly discouraged or prohibited in high-stakes contexts. Algorithmic fairness metrics should be revisited while inclusion must not mean assimilation into a flawed ontology. Again, AI systems should be built to engage with users, not to categorise them. Policymakers must also move beyond performance benchmarks and introduce requirements for ontological transparency. The future of AI must be shaped not only by technical innovation but by ontological imagination, that is, a willingness to build systems that reflect the complexity, fluidity, and multiplicity of the world we live in. Only then can artificial intelligence cease to reproduce the past and begin to participate in a more just and inclusive future.

Bibliography

- Archibong, Emmanuel I. (2024) *The Ontology of the Mind's Creativity and the Language of Artificial Intelligence*, USEM Journal of Languages, Linguistics & Literature, Vol. 12.
- Buolamwini, Joy and Gebru, Timnit (2018). “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification,” *Proceedings of Machine Learning Research* 81.
- Butler, Judith (1990). *Gender Trouble: Feminism and the Subversion of Identity* New York: Routledge,

¹⁸ Sasha Costanza-Chock, *Design Justice: Community-Led Practices to Build the Worlds We Need* (Cambridge, MA: MIT Press, 2020), 95.

- Caliskan, Aylin, Bryson, Joanna J. and Narayanan, Arvind (2017). "Semantics Derived Automatically from Language Corpora Contain Human-Like Biases," *Science* 356, no. 6334.
- Costanza-Chock, Sasha (2020). *Design Justice: Community-Led Practices to Build the Worlds We Need* Cambridge, MA: MIT Press,
- Crawford, Kate and Paglen, Trevor (2019). "Excavating AI: The Politics of Images in Machine Learning Training Sets," *Excavating AI*,
- Dastin, Jeffrey (2018). "Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women," *Reuters*, October 10,
- Fricker, Miranda (2007). *Epistemic Injustice: Power and the Ethics of Knowing* Oxford: Oxford University Press.
- Gruber, Thomas R. (1993). "A Translation Approach to Portable Ontology Specifications," *Knowledge Acquisition* 5, no. 2.
- Noble, Safiya Umoja (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism* New York: NYU Press.
- O'Neil, Cathy (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* New York: Crown.
- Searle, John (1980). "Minds, Brains, and Programs," *Behavioural and Brain Sciences* 3, no. 3 419.
- Uduk, Happiness E. and Archibong, Emmanuel I. (2024) *Buber's Complementarity and Gender Stereotyping* Port Harcourt: Gift Austine Company
- Wachter, Sandra and Mittelstadt, Brent (2019). "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI," *Columbia Business Law Review*, no. 2.