Journal on Interactive Systems, 2026, 17:1
ISSN: 2763-7719 • doi: 10.5753/jis.2026.7084

Research Paper

Towards Design Justice and Gender Perspectives - An Inclusive Audit of Facial Recognition Systems

Aildson Ferreira [Federal Rural University of Pernambuco | *aildson.ferreira@ufrpe.br*]
George Valença [Federal Rural University of Pernambuco | *george.valenca@ufrpe.br*]

Abstract. This study presents a qualitative audit of commercially available facial recognition tools featuring Automated Gender Recognition (AGR), specifically Amazon Rekognition, Face++, and Google's Vision AI, to investigate their inherent intersectional biases and broader ethical and social implications. While Artificial Intelligence offers efficiency, evaluating AI bias using isolated demographic categories overlooks intersectional discrimination, allowing deeper systemic inequalities to persist, disproportionately harming marginalized groups such as Black, transgender, and non-binary individuals. To counteract these biases, we evaluate the tools across four critical dimensions: legal and corporate responsibility, ethical considerations, social implications, and Justice, Equity, Diversity, and Inclusion (JEDI). Our findings reveal a mixed landscape. Google demonstrated a proactive shift toward algorithmic justice by removing gender labels from its Vision AI, acknowledging the problematic nature of inferring gender from appearance. Conversely, Amazon Rekognition and Face++ exhibit a concerning lack of transparency regarding their training datasets and misalignment with inclusive best practices, particularly concerning non-binary gender views and the necessity of gender identification. Ultimately, the findings reinforce that achieving algorithmic justice requires moving beyond mere inclusion to ensure user autonomy, privacy, and systemic change. Relying solely on technical accuracy is insufficient; early integration of frameworks like Design Justice and IEEE Ethically Aligned Design is essential. Furthermore, policy must evolve to prevent AI from being constrained by binary legal definitions. Establishing ethical, human-centric technology demands strong interdisciplinary collaboration to protect vulnerable populations and ensure algorithms genuinely respect the diversity of the communities they serve.

Keywords: Facial Recognition, Automatic Gender Recognition, Design Justice, Algorithmic Bias, Inclusion

Edited by: Emanuel Felipe Duarte | **Received:** 02 December 2025 • **Accepted:** 22 June 2026 • **Published:** 14 July 2026

1 Introduction

Facial recognition technologies represent a significant advancement in artificial intelligence (AI), offering features like photo organization, device unlocking, diagnosis, and surveillance. Despite the common belief that algorithms are objective and neutral, they are fundamentally imbued with biases [Silva, 2022] that arise from the initial design phases. When these biases are unchecked, AI systems can reproduce prejudices and amplify oppression on a large scale, disproportionately affecting marginalized groups [Silva, 2021]. Most facial recognition systems operate with 2D images, analyzing faces based on their geometry. They extract unique and common features such as distance between eyes, facial contours, lip size, hair length, and even clothing [Adjabi *et al.*, 2020]. When it comes to gender, algorithms often rely on outdated stereotypes [Duan *et al.*, 2025] to classify the supposed facial features attributed to each gender. Although humans more easily understand the nuances of gender expression, algorithms are not designed to capture that level of contextual and expressive subtlety.

The systemic invalidation of gender by these systems causes significant psychological harm, including damage to self-esteem, deterioration of mental health, and increased social stigma. This is often perceived as more damaging than being misrecognized by a human, as technology does not allow easy correction [Beretta *et al.*, 2024].

Social and digital marginalization is a direct consequence, creating difficulties in accessing gender-segregated spaces like restrooms and changing rooms, and essential services such as border control and access to benefits. This attacks the autonomy of transgender and non-binary individuals and their access to basic rights.

To address this issue, we conducted an audit of three gender recognition tools: *Amazon Rekognition*[1], *Vision AI*[2] and *Face++*[3]. These representative and widely-adopted facial recognition systems are assessed based on a catalog of best practices to develop automatic gender recognition solutions identified in the literature [Perilo *et al.*, 2024]. The catalog involves 19 practices divided in 4 dimensions: *legal and corporate responsibility*, *ethical considerations*, *social implications*, and *Justice, Equity, Diversity, and Inclusion in software development*.

Our results revealed that, despite several ethical concerns and persistent issues with bias, the companies developing these tools are gradually adopting transparent practices, which leads to outcomes that prioritize human well-being, such as the removal of gender recognition features. We discuss these findings considering frameworks such as Design Justice [Costanza-Chock, 2020], which are strategies to (i) address the designers' lack awareness about the social implications of their work and (ii) redefine design processes to

[1]Amazon Rekognition is available at `https://aws.amazon.com/pt/rekognition/`. Accessed: 10 July 2026.
[2]Google Cloud Vision AI is available at `https://cloud.google.com/vision`. Accessed: 10 July 2026.
[3]Face++ is available at `https://www.faceplusplus.com`. Accessed: 10 July 2026.

explicitly center who is affected by the design outcomes and could be integrated early in the development process, starting from problem formulation.

We believe that this research provides important evidence related to the protection of vulnerable communities by software companies, which also must involve strong interdisciplinary collaboration in academia. Mainly, we view this study as an example of an approach that can counter the perpetuation of structural inequality and contribute to more authentic algorithmic justice.

2 Conceptual Background

2.1 Justice and Fairness in Algorithms

The concept of **algorithmic justice** is centered on designing software solutions that embed equity, transparency, and accountability. Its core purpose is to identify and remedy biases and the exclusion of groups within machine learning (ML)-based applications. While increasingly prevalent in operational decision-making, ML algorithms can inadvertently lead to systemic discrimination against certain groups or individuals. This can result in discriminatory and unjust decisions, carrying serious consequences for marginalized and disadvantaged communities [Turchi *et al*., 2024].

This bias emerges from the problem formulation phase or from input data that already contains inherent prejudices [Coston *et al*., 2023]. Challenges in designing fair algorithms include researchers and designers potentially lacking full awareness of the social justice implications of their work, or not knowing how to incorporate mitigation strategies into the design process [Turchi *et al*., 2024]. Algorithmic bias is a significant and recognized issue in ML applications, with organizations like the IEEE and ISO actively developing standards [Turchi *et al*., 2024] to address it.

Those standards embrace the notion of **fairness metrics**, which are quantitative measures designed to capture and evaluate the impartiality or equity of a decision-making system or process. For instance, metrics related to the idea of "equal opportunities" focuses on ensuring that the true positive rates are equal across different protected groups. It emphasizes fairness in the allocation of positive outcomes [Leben, 2025]. They play a crucial role in ensuring that Machine Learning and Artificial Intelligence systems operate justly and ethically, mitigating biases that could lead to discriminatory outcomes. ML systems can generate systematic errors, known as algorithmic bias, which affect outcome predictions. Fairness metrics identify and quantify the extent of such unfairness across different population subgroups [Raza *et al*., 2024].

Among their primary applications, fairness metrics are used to identify and quantify biases, analyzing the degree of injustice among population subgroups. They are also vital during the algorithm development process, where they promote data quality and diversity by incorporating the cultural nuances of various populations. Their implementation helps ensure that AI system decisions don't result in imbalanced or prejudiced outcomes for specific groups based on sensitive characteristics like race, gender, or other social or health-related issues [Moon and Ahn, 2025; Raza *et al*., 2024]. Such metrics turn algorithmic justice into a measurable and implementable concept. Without them, it would be feasible to verify if a machine learning system operates with accuracy or efficiency, but not if it works fairly.

2.2 Gender Identity Conceptions

In simple terms, gender is not the same as anatomy; it is the way a society organizes expectations, roles, identities, and power around what it understands as "masculine" and "feminine." In the social sciences tradition, Joan Scott defines gender as a category that helps explain social relations and inequality, not just biological differences [Scott, 1986].

According to a Brazilian study from UNESP, **cisgender** refers to a person whose gender identity matches the sex and/or gender assigned at birth; **transgender** is an umbrella term for people whose gender identity does not match that initial assignment [Spizzirri *et al*., 2021]. **Non-binary** describes identities that do not fit only into "man" or "woman", or that exist between those categories or outside them. In the same study cited above, this is treated as a distinct gender identity, not as a phase or a mere preference [Spizzirri *et al*., 2021].

Misgendering is when someone is referred to using the wrong gender, for example, through pronouns, names, or markers that do not correspond to their gender identity. Even when the term itself does not always appear as a technical expression in Brazilian texts, the issue is the same discussed in research on social name, gender identity, and respect for trans people. Using the wrong form of address or denying how a person identifies produces embarrassment and exclusion [Silva *et al*., 2017; Rafael *et al*., 2023].

2.3 Design Justice

To examine how the design of social-technical systems distributes benefits and burdens, the designer and professor of MIT Sasha Constanza-Chock created the analytical and community of practice framework entitled **Design Justice** [Costanza-Chock, 2020]. It considers systems of oppression such as white supremacy, heteropatriarchy, and capitalism within technological usages.

Complementing this, ethical design methodologies, particularly regarding Autonomous and Intelligent Systems (A/IS), need approaches prioritizing human well-being, human rights, and transparency over purely functional or economic metrics. The integration of these frameworks into the development life-cycle is critical to transition from generalist design assumptions to practices that ensure accountability and the meaningful inclusion of stakeholders who are most directly impacted by technological outcomes. When intelligent systems are misaligned with these ethical strictures, they frequently function as mechanisms of discriminatory design, cis-normative assumptions automate and amplify existing social-technical disparities, inflicting disproportionate harm upon marginalized groups through systematic exclusion and erasure.

2.4 Gender Recognition by Algorithms

As a subfield of facial recognition technology, **Automatic Gender Recognition** (AGR) aims to algorithmically identify an individual's gender from photographs or videos. This technology extracts and analyzes facial, corporal, or vocal characteristics, such as bone structure, hair patterns, skin texture, or voice in a pipeline divided into four steps, which

could be described by some previous studies: **dataset** [Santos *et al*., 2023]; **training** [Soldatelli, 2022]; **binary classification** [Buolamwini and Gebru, 2018]; **inference** [Soldatelli, 2022].

- **Dataset:** The process begins with the construction of a labeled image dataset. This dataset provides the empirical basis for the model by associating each image with a target class, allowing the system to learn from examples rather than from predefined rules. Its composition, balance, and annotation quality directly influence the behavior of the entire pipeline.
- **Training:** During training, the model iteratively processes the labeled images and adjusts its internal parameters to reduce prediction error. Through this optimization process, it learns statistical regularities and visual correlations that distinguish one class from another. Training therefore transforms raw data into a predictive representation that can later be applied to unseen inputs.
- **Binary Classification:** Once trained, the model performs binary classification by mapping each input image to one of two possible categories. This stage implements the learned representation into a discrete decision, typically based on a probability score or decision threshold. In gender recognition systems, this binary output reflects how the model interprets the visual patterns it has learned during training.
- **Inference:** Inference is the deployment stage in which the trained model is applied to previously unseen images. At this point, the system no longer learns; instead, it uses the parameters obtained during training to generate predictions for new inputs. The reliability of inference depends on how well the model generalized from the training data and how similar the new image is to the data used during development.

Pipeline logic: These stages form a sequential pipeline (as described in Figure 1), the dataset provides the examples, training extracts patterns from those examples, binary classification converts those patterns into a two-class decision, and inference applies the learned decision rule to new cases. In other words, each stage depends on the quality and assumptions of the preceding one, making the pipeline cumulative rather than independent.

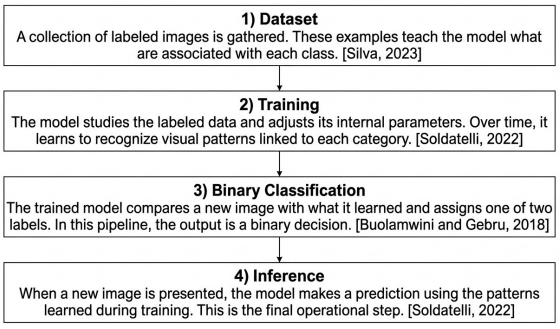

Figure 1. A brief description of how AGR technology works

Proposed applications for AGR include physical access control, data analytics, personalized advertising, and research in human-computer interaction (HCI) [Keyes, 2018]. Despite these applications, the impacts of AGR are disproportionately negative for transgender and non-binary individuals, as the technology fundamentally fails to recognize the subjectivity and fluidity of gender identity [Keyes, 2018]. This leads to a high rate of misclassifications (misgendering), which is perceived as harmful, sometimes more severe than human misgendering, and can result in significant psychological, emotional, and physical distress [Hamidi *et al*., 2018].

AGR reinforces archaic and binary gender norms, contributing to the exclusion and invalidation of non-binary identities [Keyes, 2018]. There are serious concerns regarding user autonomy, privacy, and security, as AGR can be employed for non-consensual surveillance, data collection without consent, and discrimination [Quinan and Hunt, 2022]. The technology exhibits significant limitations and algorithmic biases, often showing lower accuracy for individuals with darker skin tones or non-cis-normative facial features, thereby perpetuating existing societal prejudices [Hamidi *et al*., 2018].

Transgender and non-binary individuals are frequently utilized as "challenge sets" to refine the accuracy of AGR algorithms, which raises profound ethical questions concerning value extraction and consent [Quinan and Hunt, 2022]. Fundamentally, AGR is considered problematic because it is predicated on the notion that gender is "assigned" externally rather than being a matter of "self-identification", a premise that effectively denies the existence and lived experience of transgender individuals [Hamidi *et al*., 2018].

2.5 Inclusive AGR Solutions

By analyzing the literature on AI and software engineering, we could identify guidelines for developing more ethical solutions, which aims to ensure a responsible and inclusive approach throughout the entire life-cycle of these technologies. Those practices, which emphasize human well-being and fundamental rights, are divided in dimensions that combine legal, ethical, social, and justice-based perspectives.

Legal and corporate responsibility is rooted in governmental regulations (such as data protection laws like Brazil's LGPD) and corporate governance frameworks. It mandates transparency, corporate accountability, and strict internal policies governing biometric data collection, ensuring that individuals' fundamental rights to privacy are protected through robust data governance and licensing agreements.

Ethical considerations stem from academic discourses in human-computer interaction (HCI) and AI ethics, focusing on securing explicit, revocable user consent, privacy, trust, contestability, and user autonomy. This dimension ensures that systems respect an individual's right to self-identify and manage their own data, empowering users to contest opaque algorithmic decisions.

Social implications, heavily influenced by queer theory and the social sciences, question the foundational need for gender classification in IT systems. This dimension rejects the idea that gender is a purely binary, immutable construct that can be inferred merely from external physical characteristics, warning that such biological reductionism perpetuates harmful societal stereotypes and cis-heteronormative norms.

JEDI (Justice, Equity, Diversity, and Inclusion) originates from social justice movements and structural frame-

works like Design Justice, emphasizing the creation of diverse datasets that reflect a multiplicity of races and non-binary gender markers. It demands rigorous documentation, algorithmic fairness reporting, and system testing on non-target audiences to actively review or deprecate biased gender information.

Together, these dimensions guide a critical assessment of AGR technologies. We list the practices as follows.

- **BP1** - Analyse the context in which the system will be used, considering potential implications and risks;
- **BP2** - Create policies and standards for how gender is considered by automatic gender recognition solutions;
- **BP3** - Create policies and auditing processes to evaluate how automatic gender recognition solutions are developed and used;
- **BP4** - Ensure compliance with data protection legislation to regulate the collection and processing of biometric data;
- **BP5** - Ensure corporate accountability and due diligence;
- **BP6** - Create licensing agreements for database use;
- **BP7** - Ensure data governance;
- **BP8** - Ensure diversity and inclusion within design and development teams;
- **BP9** - Promote awareness initiatives to discuss gender, bias and IT development;
- **BP10** - Ensure that users consent for implementation;
- **BP11** - Ensure safety, privacy, transparency, trust, contestability and autonomy for users;
- **BP12** - Assess the need of gender identification;
- **BP13** - Ensure that not only external characteristics are considered to infer a person's gender;
- **BP14** - Include a non-binary view of gender;
- **BP15** - Create inclusive datasets, with diversity in terms of race, gender and/or markers;
- **BP16** - Revisit, revise, or retract existing image databases, also collecting gender information posted online;
- **BP17** - Ensure database transparency and documentation;
- **BP18** - Conduct system testing on non-target audience;
- **BP19** - Create reports about algorithms' performance and fairness metrics.

These practices were developed via a rigorous academic and empirical synthesis by [Perilo *et al*., 2024]. They conducted a systematic mapping of AI literature and validated these guidelines through interviews and focus groups involving both technology experts (such as AI developers and data scientists) and individuals directly affected by these technologies, including trans and non-binary community managers and the CEO of a trans employment agency. The main goal is to analyze the use context of AGR systems, comprehending the significance of gender data and the potential impact of derived decisions (particularly in public policy contexts). To achieve this, it is crucial to *de-biologize* [Oyěwùmí, 2021] and expand the concept of gender, acknowledging the diversity of identities and rejecting purely binary classifications.

However, the necessity of these guidelines invites a critical reflection on the current state of technology: *what happens when marginalized groups are historically absent from the actual co-creation process?* When transgender and non-binary individuals are excluded from problem formulation and the early design phases, AI systems inevitably reproduce and amplify structural oppressions, such as cis-normativity, white supremacy, and hetero-patriarchy. This systemic exclusion is profoundly problematic from both a social and political perspective. Politically, it allows unregulated, biased technologies to dictate access to fundamental rights, physical spaces (like borders and restrooms), and essential services, effectively institutionalizing algorithmic discrimination and mass surveillance. Socially, it invalidates the lived realities of marginalized populations, causing severe psychological harm, reinforcing stigma, and stripping individuals of their sovereignty over their own identities.

Therefore, the present research adopts this catalog[4], as it collectively works to prevent AI from reinforcing societal prejudices and to foster technological development that is truly just and equitable for transgender and non-binary individuals.

3 Research Method

This study involved a document analysis to understand complex problems and phenomena related to algorithmic justice. Hence, two researchers reviewed and interpreted varied information in pre-existing artifacts from the three facial recognition systems selected. Figure 2 presents the three main phases of our study, which we describe in the subsequent sections.

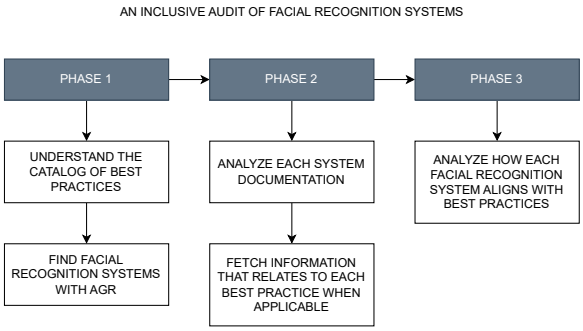

Figure 2. Research phases of our methodology, from data gathering to analysis of findings.

3.1 Phase 1 - Tools Selection

We started this phase by understanding best practices for developing automatic gender recognition solutions. Our goal was to define a set of distinct facial recognition tools and obtain their public documentation, which we would evaluate according to the practices. Then, to start the data collection, we identified facial recognition tools via online search engines. Initial searches focused on keywords such as "*facial recognition tools*", "*facial recognition APIs*" and "*computer vision systems*". The preliminary and most relevant results consistently indicated the following platforms: *Amazon Rekognition*, *Face++, Vision AI*, *Microsoft Azure Face API* and *IBM Watson Visual Recognition*. The relevance and recurrence of these tools in the initial searches established them as potential candidates for analysis in this study. To select facial recognition tools for analysis, we adopted a process based on predefined

[4]The complete catalog of best practices is available at [Perilo *et al*., 2024]

inclusion (IC) and exclusion (EC) criteria, as detailed in Table 1.

Table 1. Criteria to select facial recognition systems.

ID	Criteria
IC1	Explicit capability to infer gender from faces in 2D images or videos
IC2	Availability of comprehensive documentation written in English
IC3	Availability for commercial use via APIs and SDKs
EC1	Absence of gender recognition functionality
EC2	Absence of documentation written in English

In parallel, we used digital engines, such as the ACM Digital Library and IEEE Xplore, adopting a specific search string (Table 2). We aimed to identify studies addressing biases in facial recognition systems, with a particular focus on the LGBTI+ community.

Table 2. Search string to find previous studies the topic.

Search String
`(("Document Title":facial recognition OR "Document Title":gender recognition AND "Document Title":analysis OR "Document Title":audit OR "Document Title":design) AND ("Full Text Only":bias) AND ("Full Text Only":rekognition OR "Full Text Only":face++ OR "Full Text Only":google OR "Full Text Only":ibm OR "Full Text Only":microsoft) AND ("Full Text Only":lgbt OR "Full Text Only":trans OR "Full Text Only":non-binary))`

We found a significant gap in the scientific literature concerning these biases for the community. Given the scarcity of specific works, the selection of tools for subsequent analysis was guided by their market prominence, the controversies and public stances surrounding them, especially concerning issues of algorithmic justice and demographic discrimination. Hence, the following three were selected: Amazon Rekognition, Vision AI, and Face++.

Recently, the Chinese tool Face++ was notably analyzed in the Gender Shades study [Buolamwini and Gebru, 2018], which revealed not only racist biases in its face detection but also exceptionally high error rates in gender prediction for Black women compared to White men. In their turn, Amazon Rekognition and Vision AI, both North American, were selected for comparative purposes due to their operation within the same segment of automated gender recognition, despite also having documented issues in facial recognition. We describe each of these tools in the following paragraphs.

Amazon Rekognition is a fully managed computer vision service that enables the integration of image and video analysis into applications without requiring extensive machine learning expertise. It offers comprehensive features for identity verification, including face detection and comparison, facial search, and liveness detection to prevent fraud. Additionally, the service assists in content moderation by identifying unsafe elements, facilitates the creation of searchable media libraries through object and scene detection, and performs celebrity recognition. It can detect personal protective equipment and events in videos, and also allows for the creation of custom labels for specific business objects and detects text in images. All these functionalities are easily integrated via APIs and are compatible with other AWS services [Amazon, 2025]. While many large-scale implementations are proprietary, available information indicates the tool's application across various sectors. For example, a sheriff's department in Oregon has utilized the service to compare individuals' faces against a database of mugshot photos [Union, 2018].

Despite its capabilities, Amazon Rekognition faces significant controversies related to its accuracy and the ethical implications of its use. Tests conducted by the ACLU (American Civil Liberties Union) in 2018 revealed that it incorrectly identified 28 members of Congress with mugshot photos, exhibiting a disproportionate racial bias. Nearly 40% of these false matches involved people of color, despite this demographic representing only 20% of Congress [Union, 2018].

Amazon defended itself by claiming misuse of the software, asserting that the ACLU did not adhere to the recommended 99% confidence threshold for security applications. However, this defense raises critical questions regarding responsibility in AI deployment. It suggests that if a system produces significant errors in its default configurations or with common operational practices in sensitive applications, the burden of accuracy should not solely fall on the end-user. Such separation between developer recommendations and operational realities, particularly in high-risk domains like law enforcement, may cause critical errors.

Vision AI classifies Hispanic or Black individuals as potential scientists less frequently than White individuals [Mohammadi *et al.*, 2025]. Vision AI is a platform leveraging computer vision and artificial intelligence to extract information from images, documents, and videos, offering both pre-trained and customizable models [Google, 2025b]. Its capabilities encompass general visual content processing, including object detection and recognition, image classification and search, and the ability to generate captions and respond to multimodal queries. For document understanding and processing (Document AI), the tool provides advanced Optical Character Recognition, data extraction, and workflow automation. The Video Intelligence API facilitates object detection and tracking in videos, scene understanding, and content moderation. In industrial and manufacturing applications (Visual Inspection AI), it automates visual inspection and defect detection [Google, 2025b].

The tools within Vision AI are used in a variety of applications. In the retail sector, the Cloud Vision API automates product labeling. Document AI processes large volumes of documents in paper-intensive industries. Furthermore, Visual Inspection AI automates quality inspection in manufacturing, effectively detecting anomalies and defects [Marangon, 2025]. In the financial sector, Banco BV leverages Google Agentspace (which integrates Gemini and Google Search) to empower its employees with AI agents, optimizing discovery and automation within their systems [Grannis, 2025]. Deloitte also employs Agentspace to connect data sources, drive experimentation, and identify relationships within reports. Finally, the Video Intelligence API is applied in content moderation and recommendation, media creation and contextual

ad insertion [Grannis, 2025].

However, Vision AI faces significant controversies and ethical concerns, primarily related to algorithmic bias. Incidents such as the mislabeling of African Americans as "gorilla" in 2015 [Kayser-Bril, 2020] and the misidentification of a dark-skinned individual with a thermometer as "weapon" in 2020 [Kayser-Bril, 2020] revealed persistent racial biases which are often attributed to unrepresentative training data, inherent algorithmic design flaws, and a lack of diversity within development teams. Such failures denote the perpetuation of stereotypes, unequal access to information, and potential psychological impacts on vulnerable communities, in addition to the proliferation of AI-generated misinformation and deepfakes [ScienceDaily, 2025], which represent serious risks.

Face++ is a facial recognition platform that offers high-precision and high-speed facial detection, recognition, and analysis, including crucial liveness detection for fraud prevention [HFSecurity, 2023]. Its applications are extensive, encompassing security and surveillance, where it is employed for real-time identification and tracking in public spaces and access control, as well as assisting law enforcement agencies in forensic analysis and suspect identification. In the financial services sector, Face++ facilitates secure identity verification and biometric authentication, exemplified by Alipay's "*Smile to Pay*" feature [HFSecurity, 2023]. Within healthcare, it optimizes patient identification, aids in the diagnosis of genetic disorders and mental health issues, and enhances hospital security. The technology is also adopted in human resources for attendance tracking and candidate verification, and in entertainment, gaming, social media, and photography for AR filters, avatar creation, and photo editing [HFSecurity, 2023].

The deployment of facial recognition systems faces significant criticism due to documented instances of privacy violations, erroneous identifications, and overarching ethical concerns. Face++ (Megvii) exhibits inherent precision limitations influenced by factors like lighting and facial expressions. It demonstrates higher error rates for Black individuals and females from 18 to 30 years of age [Buolamwini and Gebru, 2018]. The U.S. Department of Commerce even placed Face++ on an "exclude list" because of alleged human rights violations against Muslim minorities in Xinjiang, which links the technology to state oppression [Paganoni *et al*., 2019].

In the healthcare sector, the use of facial recognition raises questions concerning informed consent, the risk of biased diagnoses stemming from non-diverse training data, the privacy of sensitive biometric data, and the potential erosion of patient-clinician trust [Martinez-Martin, 2019]. These challenges require more rigorous ethical and regulatory scrutiny, particularly in sensitive domains (e.g., healthcare), to ensure accountability and transparency where no specific laws protect transgender and non-binary individuals.

3.2 Phase 2 - Data Extraction

The data collection involved reviewing the official documentation provided by the companies maintaining the selected tools. For each of practice, a systematic search was conducted within each tool's official documentation. In cases where the information was nonexistent or insufficient for a clear evaluation, the research was expanded to traditional media (informative websites) to locate data that could supplement or contextualize how the practices are adopted by these tools. Our primary goal was to map the explicit information that either claims or refutes adherence to a specific practice (as shown in Figure 3). This allowed us to evaluate the compliance of the company with the catalog[5].

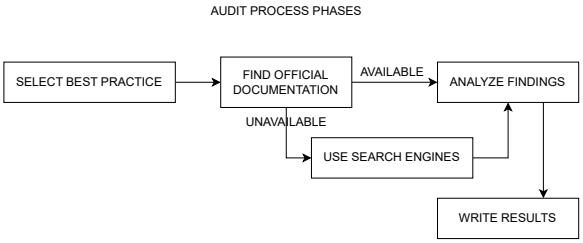

Figure 3. Activities to Analyze Facial Recognition Systems.

3.3 Phase 3 - Synthesis of Findings

This phase focused on evaluating the extracted data to determine each company's compliance with a defined catalog of practices. To achieve this, we searched for the presence of documentation addressing each practice. A "*follows*" classification, indicating compliance, was assigned when such information existed. Conversely, a "*does not follow*" classification was given in the absence of such explicit mention in the official documentation or related information, regardless of the tool's alignment to the practice. This approach ensures the evaluation is based exclusively on the availability and clarity of publicly accessible information. It reflects the transparency and maturity of the tool's documentation as a key indicator of its adherence to established best practices. The information gathered from both official documentation and supplementary sources was then synthesized into a comprehensive compliance table, which allowed us to identify clear patterns and discrepancies. The final synthesis provides a clear overview of the landscape concerning how these tools align with ethical AI development.

3.4 Limitations and Methodological Positioning

Documentation serves as the primary, authoritative source of truth for a tool's intended functionality, limitations, and design assumptions, enabling a systematic analysis of explicit claims, such as binary gender classifications or the absence of non-binary options, without the variables introduced by real world testing, like API rate limits, evolving model updates, or undocumented edge cases. According to [Scheuerman *et al*., 2019], this approach aligns with established practices in AI ethics auditing, where preliminary evaluations often prioritize declarative transparency over black-box empiricism, as evidenced by studies critiquing commercial face recognition services for lacking non-binary support directly from their APIs and terms. Moreover, it upholds empirical rigor by mitigating biases in empirical validation (dataset scarcity for trans and non-binary identities), focusing instead on verifiable developer disclosures that reveal systemic gaps in inclusion, such as incomplete gender lens or unstated performance metrics issues that persist even if empirical tests were feasible but

[5]The detailed audit results are available at `https://doi.org/10.5281/zenodo.21302252`

would still be gated by access barriers and ethical concerns over data usage. In fields like AI bias research, this methodology is defended as "*minimum necessary rigor*" for scoping studies, offering a transparent baseline for future empirical inquiry that flags limitations and avoids over-interpreting causal results [Klein *et al*., 2023].

4 Results

The auditing process revealed a significant lack of institutional transparency regarding the internal mechanisms and training datasets of commercial Automated Gender Recognition (AGR) tools. While providers often emphasize a commitment to responsible AI through high-level principle documents, there is a discrepancy between these public ethical frameworks and the granular technical documentation available to researchers.

This opacity presented methodological challenges during the construction of this analysis. The investigation was primarily obstructed by the black-box nature of proprietary environments, which prevented the verification of demographic provenance, consent protocols, and metadata in training datasets. In addition, audit efforts were restricted by documentation that frequently relies on marketing terminology or generic sustainability reports rather than providing standardized technical metrics for fairness. The persistence of rigid binary classification in tools like Amazon Rekognition and Face++, despite academic consensus on its limitations, suggests a prioritization of commercial statistical utility over the accurate representation of diverse gender identities.

The absence of standardized Model Cards or detailed metadata forced this analysis to rely on indirect empirical evidence and external audits to verify corporate claims of fairness. This underscores a critical need for required transparency in the AI industry to allow for independent scientific validation of ethical compliance.

To defend whether a company complies with or fails to follow a best practice, we rely on the presence of explicit, publicly accessible documentation or unofficial evidence. Otherwise, it is classified non-compliant regardless of its internal operations. For Legal and Corporate Responsibility (*BP1-BP9*), compliance requires public records of comprehensive risk assessments, strict data governance frameworks, and explicit alignment with data protection laws, alongside policies that define gender beyond mere biology and prove genuine transgender representation in design teams. In terms of Ethical Considerations (*BP10-BP11*), a company must prove it secures clear, revocable user consent without complex jargon and provides mechanisms for users to autonomously access, contest, or opt out of system categorizations. For Social Implications (*BP12-BP14*), compliance is defended by publicly questioning the actual necessity of gender identification in the system, explicitly refusing to infer gender solely from external physical characteristics, and formally embracing a fluid, non-binary perspective. For JEDI guidelines (*BP15-BP19*) companies must have documented proof that training datasets are richly diverse and inclusive of marginalized gender markers, coupled with clear efforts to retract or revise biased legacy databases. Furthermore, companies must also provide complete transparency regarding data composition and publish fairness reports detailing rigorous system testing on non-target audiences to proactively catch and mitigate algorithmic disparities.

In Table 3 we present the results, denoted as "*follows*", which is represented by a *checkmark* (✓). Conversely, non-compliance, or "*does not follow*", is indicated by the absence of any mark. This binary notation enhances clarity and facilitates rapid assessment of adherence to best practices.

Table 3. Comparative analysis of three different computer vision tools alignment with Best Practices (BP) grouped by dimension.

BP / Tool	Amazon Rekognition	Face++	Vision AI
Legal and corporate responsibility			
BP1			✓
BP2			
BP3			✓
BP4	✓	✓	✓
BP5	✓		✓
BP6	✓		✓
BP7	✓		✓
BP8			✓
BP9	✓		✓
Ethical considerations			
BP10	✓	✓	✓
BP11	✓		✓
Social Implications			
BP12			✓
BP13			
BP14			✓
JEDI (Justice, Equity, Diversity and Inclusion)			
BP15			✓
BP16			
BP17			✓
BP18			✓
BP19			✓

Amazon Rekognition. The results show that Amazon Rekognition performs gender prediction based on physical facial characteristics. AWS documentation clarifies that this detection isn't intended to categorize gender identity but rather to assist in statistical analyses through individual specificities. These characteristics are used to create metadata from facial analysis, and the documentation even cites a case of a dating app in India that used Rekognition to analyze profiles to ensure the company's "high standards" [Amazon, 2019].

Although tests on Rekognition technology in 2018 demonstrated high false-positive rates for individuals with darker skin, Amazon does not provide public data specifying the training datasets used for Rekognition's facial recognition models. Instead, it transfers the responsibility of ensuring data adequacy to the service's users themselves. While there are no specific public policies on how gender should be considered by the tool, Amazon does offer Amazon SageMaker Clarify, an AWS service that helps detect biases and explain predictions, contributing to model review and the reduction of classification disparities.

The data concerning Rekognition reveals a stance that, in part, addresses legal and corporate responsibility but also presents some gaps. The public documentation doesn't detail

an explicit risk analysis of the results, although it recommends human oversight for potential false positives. A point of concern arises when Amazon discontinues outdated programs and materials, including DEI (Diversity, Equity, and Inclusion) initiatives [Reuters, 2025; McLymore, 2025]. This decision sharply contrasts with the company's role in discussing responsible and secure AI, especially when Amazon meets with the U.S. government and industry leaders to discuss the responsible and safe use of AI [Amazon, 2023], demonstrating an apparent disconnect between its external public stances and its internal diversity and inclusion policies, which are fundamental to mitigating algorithmic biases.

Face++. The analysis of Face++ reveals a significant lack of public transparency, contrasting sharply with growing expectations for corporate responsibility within the technology sector. Although Megvii, as a major AI company and the developer of the tool, possesses internal policies and agreements regarding the data it acquires for model training, the specifics of these licensing contracts are proprietary and not publicly disclosed.

The absence of documentation directly addressing the risks associated with its results is particularly concerning, especially given the sensitive nature and ethical implications of facial recognition and the controversies surrounding the tool. Furthermore, the non-existence of public data governance policies and the lack of information regarding gender diversity initiatives within design teams to mitigate biases suggest that the company either does not publicly address the root causes of algorithmic biases in its products or does not treat it as a critical issue. In addition, Face++ operates under Chinese data protection laws, which are evolving (such as the *Personal Information Protection Law* - PIPL). However, these laws and their enforcement differ significantly from regulations like Europe's GDPR, with international concerns often focusing on state access to and the use of biometric data by Chinese companies.

Vision AI. While Google does not publicly detail direct risk analyses of results or explicit internal audit processes for its own Automatic Gender Recognition (AGR) systems, it demonstrates a proactive engagement with transparency and bias mitigation via other initiatives. The existence of programs like the "Inclusion Champion Program" attempts to integrate diversity into the development of its products.

Vision AI explicitly aligns with the Google AI Principles [Google, 2018], which emphasize responsible development, human oversight, and the prioritization of human well-being. Google removed gender classification from its Cloud Vision API in February 2020 after recognizing that gender cannot be reliably inferred from appearance alone and that such predictions can amplify bias and misclassification, especially for trans and gender-nonconforming people and since then, the system classifies individuals as "*person*" rather than "*man*" or "*woman*". In computer vision, this decision possibly marks an important shift toward acknowledging the limits of visual inference and rejecting the assumption that models can neutrally "read" socially constructed identities from images. From a societal perspective, it reduced the risk of automated misgendering, stereotype reinforcement and the expansion of identity-based surveillance. In terms of algorithmic fairness and design justice, the removal represents a meaningful intervention because it avoids a harmful binary classification task rather than attempting to optimize it, while also aligning system design more closely with human dignity, inclusion, and the principle that not all technically feasible inferences should be deployed.

The Google AI Principles also serve to promote transparency and corporate responsibility. Google stands out for its conscious and responsible approach in the field of artificial intelligence. Its research aids in the visualization of biased data, seeking to positively influence the technology community, as exemplified by the aforementioned removal of the automatic gender recognition feature.

In the following sections, we interpret our analysis based on the dimensions (D) of the catalog.

4.1 D1 - Legal and Corporate Responsibility

We observed that the three tools exhibit distinct approaches, ranging from complete data transparency to active and at times controversial engagement with corporate accountability:

- **Face++** demonstrates the most significant gap in public data transparency, with a near-total absence of accessible documentation on risk analysis, bias audits, or ethical governance. Its operation under a Chinese legal framework further amplifies concerns regarding state access to biometric data.
- In contrast, **Amazon Rekognition** and **Google Vision** show greater effort in communicating their policies, although with important nuances. While Amazon Rekognition recommends human oversight and employs explicit consent terms, its decision to discontinue internal DEI programs creates an inconsistency regarding bias mitigation, which is fundamental for corporate responsibility in AI technologies.
- However, despite the lack of public details about its risk analyses or specific internal audits for AGR, **Vision AI** stands out due to the robustness of its data governance policies. These include explicit consent for data usage in model improvement, active DEI initiatives focused on product development, and contributions to the broader community through research and open-source tools designed to identify biases. This approach reflects a more diligent stance in building responsible AI.

External audits, such as the Gender Shades study [Buolamwini and Gebru, 2018] (which exposed performance biases in models from Face++, IBM, and Microsoft) underscore the critical need for transparency and independent oversight. These findings highlight that self-regulation by companies is often insufficient to ensure complete accountability, and that integrating diversity into development teams is as crucial as implementing formal policies to combat structural biases in the datasets and algorithms utilized.

4.2 D2 and D3 - Ethical and Social Implications

In the second and third dimensions, the tools exhibit varying degrees of transparency and alignment with best practices:

- **Face++** maintains a pronounced lack of transparency regarding user control over ethical considerations. It per-

sists with a binary gender approach that has demonstrably proven biases, leading to severe social implications, particularly for marginalized groups.
- **Amazon Rekognition** occupies a middle ground, showing discrepancies between its stated policies and the results of bias audits. While its documentation asserts recognition of only two genders, it simultaneously maintains celebrity databases that include non-binary gender classifications.
- **Vision AI** distinguishes itself with a more robust and proactive ethical framework. It stands out for its comprehensive consent policies, transparent data practices facilitated by public tools, and detailed AI principles. Furthermore, Google actively pursues initiatives focused on diversity and bias reduction. However, explicit details regarding end-user contestability and mechanisms for user autonomy, ensuring systems respect individual self-identification and provide users control over their data are not yet universally detailed across all its services.

4.3 D4 - Justice, Equity, Diversity, and Inclusion

The final dimension is the most critical for developers:

- Here, Vision AI reinforces its commitment to transparency and its efforts regarding data diversity in the datasets used for facial recognition systems. Google provides documentation on which data are used and how, through tools such as Model Cards [Google, 2025a]. It also contributes to the scientific community via research in the field of fairness in machine learning, which further aids in the development of more inclusive datasets.
- In direct opposition, neither **Amazon Rekognition** nor **Face++** provide data specifying the training datasets used for their facial recognition models, even after the controversies involving both tools. The companies responsible for these tools do not appear to prioritize transparency and ethical reporting of their algorithms.

5 Discussion

Up to 2024, many leading technology companies, including Google, were seen as pioneers in promoting Diversity, Equity, and Inclusion (DEI). They established ambitious workforce representation goals and invested significantly in related programs, as detailed in the Google Diversity Annual Report of 2023. This approach was often considered fundamental for driving innovation, attracting talent, and reflecting the diversity of their global user bases. These corporate commitments to DEI were further strengthened following the murder of George Floyd in May 2020 [Jeyaretnam, 2025], which ignited a widespread movement for racial justice in the U.S., leading many companies to intensify their inclusion efforts.

The 2024 U.S. presidential election, with Donald Trump's projected victory, created a new political climate that significantly influenced corporate DEI policies. The then-U.S. President's administration targeted DEI initiatives via executive orders, prohibiting their funding by government agencies, revoking previous pro-diversity orders, and requiring federal contractors to eliminate explicit diversity-based hiring goals. This presented a compliance risk for technology companies heavily reliant on government contracts [Tafesse, 2025]. Amazon figured among the companies that adjusted their diversity and inclusion initiatives, ending some programs and combining various employee groups in a single team.

Coordinated and intersectional strategies can effectively mitigate the pervasive issues associated with biased AI. Related research highlights the importance of combining technical innovation, robust data governance, and well-conceived, inclusive policies. According to a study published this year, "*The Ethics of AI at the Intersection of Transgender Identity and Neurodivergence*" [Parks, 2025], a crucial starting point involves reconsidering data collection and labeling practices to adopt more inclusive approaches that account for the diverse spectrum of gender identities and cognitive profiles. The current architecture of AI also needs rethinking to proactively identify discrepancies during the training phase, always ensuring transparency and continuous user feedback. This iterative process allows systems to evolve in alignment with community needs.

Another related work, "*Fairness Through Feedback*" [Quaresmini and Zanotti, 2025], proposes a feedback mechanism where users can correct AGR system classifications. This allows individuals to validate or modify the predicted label, reflecting the principle that gender identity is a matter of self-identification and that the user should have the final say, even if it might reduce the system's "autonomy".

Governmental policies can either support or undermine inclusive AI design. Policies that impose binary gender definitions, for example, hinder the collection and labeling of comprehensive data. It is essential that AI data processing isn't constrained by legal frameworks that deny certain forms of gender recognition or view other diversities through a deficit model. We must also ensure that marginalized communities actively participate in problem formulation, co-creation of solutions, and impact evaluation. Our findings indicate significant disparities in the practices concerning gender classification, the use of inclusive datasets, and data transparency. Specifically, both Amazon Rekognition and Face++ exhibit notable misalignment with recommended best practices as outlined in Best Practices 12 to 14, which address the assessment of gender identification necessity, the reliance on purely external characteristics for gender inference, and the inclusion of non-binary gender views. The removal of gender labels from Vision AI, for instance, reflects an evaluation that inferring gender through appearance is problematic and unnecessary due to potential bias. Decisions like these, means moving beyond mere "inclusion" to seek justice, autonomy, and sovereignty for users over their data and identities.

To achieve this, we can combine frameworks like the *IEEE Ethically Aligned Design* (EAD). EAD advocates for prioritizing human well-being and ethical considerations throughout the life-cycle of Autonomous and Intelligent Systems (A/IS). It recommends establishing governance structures, such as standards and regulatory bodies, to ensure AI respects human rights, freedom, dignity, and privacy, placing human advancement at the core of technological development. This approach fosters transparent and explainable systems that users can understand and trust.

Ethical design standards are also crucial. Best Practices 15 to 19, address the necessity of inclusive and diversified

datasets, transparency in models and documentation and bias mitigation across varied demographic groups, with a collaboration with external research and civil society for independent audits and social impact assessment. Analysis reveals a notable misalignment by Amazon Rekognition and Face++ to these guidelines: (i) Amazon Rekognition demonstrates a lack of public data regarding its training sets and delegates equity responsibility to users, while (ii) Face++ exhibits a significant deficit in data transparency and bias audits. Both systems persist in binary gender inference based on physical characteristics and stereotypes.

Eventually, applying the Design Justice framework [Costanza-Chock, 2020] to our findings underscores the urgent need to redefine AI development processes. Rather than treating marginalized groups merely as "challenge sets" for algorithm refinement, companies must explicitly center the people most affected by these design decisions—particularly those historically marginalized by the *matrix of domination* [Collins, 1990], including white supremacy, hetero-patriarchy, and ableism. Overcoming the severe ethical deficits observed in tools like Amazon Rekognition and Face++ requires moving beyond corporate self-regulation. It demands full inclusion, accountability, and control by communities with direct lived experience, ensuring that technological creation strictly adheres to the principle of "*nothing about us without us*".

6 Conclusion

Our study revealed a nuanced landscape, but one that demands urgent and radical intervention. Despite various ethical concerns and persistent issues with bias, some companies are proactively developing transparent practices, leading to outcomes that prioritize human well-being, such as the removal of gender recognition features. However, the persistence of algorithmic biases in these systems continues to pose severe threats to the fundamental rights, autonomy, and safety of transgender and non-binary individuals. The deployment of Automated Gender Recognition (or any other AI-based system) without rigorous ethical and legal guardrails is not merely a technical flaw but a structural mechanism that automates discrimination, invalidates marginalized identities, and violently dictates access to physical spaces and essential services. Politicians must establish strict legal frameworks that reject binary constraints and protect vulnerable communities from institutionalized surveillance and biometric data abuse. Simultaneously, universities and software companies must collaboratively foster inclusive design, transparency, and accountability to prevent AI from functioning as an extension of structural oppression. This ecosystem of innovation and regulation must actively center transgender and non-binary people in every phase of problem formulation, co-creation, and auditing by default. As long as these marginalized communities are treated merely as data points or challenge sets for algorithm refinement rather than active decision-makers, technology will remain a tool of marginalization.

6.1 Contribution and Related Work

Our study emphasizes the critical need for specific legislation to regulate intelligent systems, particularly concerning ethics and inclusion. As society inherently remains diverse, the algorithms designed to serve it must mirror this diversity. The results of this audit reveal significant gaps in the construction of the analyzed tools, underscoring the challenges faced by transgender and non-binary individuals when interacting with facial recognition technologies.

The comparative analysis highlights a critical divergence: while Amazon Rekognition and Face++ continue to employ binary gender classification and exhibit shortcomings in transparency and data governance, Vision AI demonstrates a proactive effort to mitigate biases by classifying individuals as "*person*" and eliminating binary gender classification. These findings contribute to a greater awareness of algorithmic biases and injustices inherent in AI tools. They reinforce the imperative for technology to be more equitable, transparent, and respectful of user autonomy and identity. This aligns seamlessly with initiatives such as the IEEE's "*Ethically Aligned Design*", advocating for the responsible development and deployment of AI systems.

In a previous study, [Reis *et al*., 2023] show that privacy policies (a critical aspect of any facial recognition technology) in Brazilian apps are often long, hard to understand, and lack objectivity, weakening user trust and transparency. Nunes et al. [Nunes *et al*., 2024] reveal that documentation tools like Model Cards foster developer "sociotechnical reflection", yet teams frequently omit ethical risks that were considered but rejected, exposing a gap between internal intent and public accountability. Our research extends these insights by empirically auditing deployed commercial vision APIs, evaluating how sociotechnical attributes (e.g., gender labeling and identity categories) are governed in practice. While the cited works diagnose limitations in documentation clarity and ethical disclosure, our goes further by assessing real system behavior and documentation maturity through a design-justice and non-binary inclusion lens, advancing from reflective analysis to measurable sociotechnical compliance and accountability in interactive AI.

6.2 Threats To Validity

A primary limitation of this study was restricted access to certain academic databases. This constraint prevented the gathering of other potentially relevant articles that could have enriched our literature review and the inability to access these works may pose a threat to the validity of our conclusions. The inclusion of additional scholarly works could have provided further theoretical context and improved our interpretation of the results, potentially revealing details not captured by our accessible sources. A primary concern is **construct validity**, particularly given the nascent and evolving nature of ethical AI best practices. The *Catalog of Best Practices* itself, while compiled diligently, may not fully encompass all nuances or future developments in ethical AI, potentially leading to an incomplete or slightly misaligned assessment of the systems' true ethical posture. In addition, the reliance on a diverse range of sources (including scientific literature, informative websites, blogs, and official documentation) may threaten **conclusion validity**, as different forms of evidence can make interpretation harder (although it support data triangulation). The varied formality and intent of these sources could lead to inconsistencies in how information is weighed or understood, potentially influencing the perceived alignment of the systems with the best practices. At last, **external validity** might be

limited due to the specific selection of only three systems for audit; while this provides in-depth analysis, the findings may not be broadly generalizable to the vast and rapidly changing landscape of facial recognition technologies from other developers or deployed in different contexts.

6.3 Future Work

Given the persistence of biases in training datasets (which have historically under-represented or excluded marginalized groups such as Black individuals, leading to disproportionately high error rates and unjust arrests) a future project will investigate the extent to which facial recognition technologies can be considered acceptable forms of biometrics. This is particularly relevant since the external inference of gender is fundamentally incompatible with the lived realities of transgender individuals and can cause significant psychological and social harm.

Additionally, we will organize a hackathon aimed at applying ethical AI development practices identified in previous studies to real-world innovation processes. This initiative, which will be guided by principles of equity, diversity and inclusion, can also raise the awareness of students and practitioners about AI systems and digital platforms' influence on human rights and well-being.

Declarations

Authors' Contributions

AF: conceptualization, data curation, formal investigation methodology, project administration, writing – original draft. GV: conceptualization, data curation, formal investigation methodology, project administration, validation, writing – original draft.

Competing interests

The authors declare that they have no competing interests

Availability of data and materials

To ensure transparency and reproducibility, the detailed audit results and compliance mapping, is openly available in Zenodo: `https://doi.org/10.5281/zenodo.21302252`.

Further relevant information

Use of Generative AI: During the preparation of this work, the authors utilized Generative AI tools to assist with translating the manuscript into English and improving its grammar and readability. After using the tool, the authors reviewed the text and take full responsibility for the final content.

Ethics Statement: This study did not involve human participants, human biological material, or personal data collection. The research consisted exclusively of documentary analysis of publicly available documentation and publicly accessible information regarding commercial AI systems. Therefore, according to institutional regulations, formal ethics committee approval and informed consent were not required.

References

Adjabi, I., Ouahabi, A., Benzaoui, A., and Taleb-Ahmed, A. (2020). Past, present, and future of face recognition: A review. *Electronics*, 9(8). DOI: https://doi.org/10.3390/electronics9081188.

Amazon (2019). Amazon rekognition launches enhanced face analysis. https://aws.amazon.com/about-aws/whats-new/2019/03/amazon-rekognition-launches-enhanced-face-analysis/, Accessed: 10 July 2026.

Amazon (2023). Our commitment to the responsible use of ai. https://www.aboutamazon.com/news/company-news/amazon-responsible-ai, Accessed: 10 July 2026.

Amazon (2025). What is amazon rekognition? https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html, Accessed: 10 July 2026.

Beretta, E., Voto, C., and Rozera, E. (2024). Decoding faces: Misalignments of gender identification in automated systems. *Journal of Responsible Technology*, 19:100089. DOI: https://doi.org/10.1016/j.jrt.2024.100089.

Buolamwini, J. and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Friedler, S. A. and Wilson, C., editors, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pages 77–91. PMLR. `http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf`, Accessed: 10 July 2026.

Collins, P. (1990). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment. Routledge.

Costanza-Chock, S. (2020). *Design Justice: Community-Led Practices to Build the Worlds We Need*. The MIT Press.

Coston, A., Kawakami, A., Zhu, H., Holstein, K., and Heidari, H. (2023). A validity perspective on evaluating the justified use of data-driven decision-making algorithms. In *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pages 690–704, Los Alamitos, CA, USA. IEEE Computer Society. DOI: https://doi.org/10.1109/SaTML54575.2023.00050.

Duan, W., Li, L., Freeman, G., and McNeese, N. (2025). A scoping review of gender stereotypes in artificial intelligence. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, pages 1–20, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3706598.3713093.

Google (2018). Our ai principles. https://ai.google/principles/, Accessed: 10 July 2026.

Google (2025a). Google's model cards. https://modelcards.withgoogle.com/model-cards, Accessed: 10 July 2026.

Google (2025b). Vision ai: Image and visual ai tools. https://cloud.google.com/vision, Accessed: 10 July 2026.

Grannis, W. (2025). 2025 and the next chapter(s) of ai. https://cloud.google.com/transform/2025-and-the-next-chapters-of-ai, Accessed: 10 July 2026.

Hamidi, F., Scheuerman, M. K., and Branham, S. M. (2018). Gender recognition or gender reductionism? the social implications of embedded gender recognition systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3173574.3173582.

HFSecurity (2023). what is face++ face recognition.

https://hfsecurity.cn/face-face-recognition/, Accessed: 10 July 2026.

Jeyaretnam, M. (2025). These u.s. companies are not ditching dei amid trump's crackdown. https://time.com/7261857/us-companies-keep-dei-initiatives-list-trump-diversity-order-crackdown/, Accessed: 10 July 2026.

Kayser-Bril, N. (2020). Google apologizes after its vision ai produced racist results. https://algorithmwatch.org/en/google-vision-racism/, Accessed: 10 July 2026.

Keyes, O. (2018). The misgendering machines: Trans/hci implications of automatic gender recognition. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW). DOI: https://doi.org/10.1145/3274357.

Klein, G., Hoffman, R. R., Clancey, W. J., Mueller, S. T., Jentsch, F., and Jalaeian, M. (2023). "minimum necessary rigor" in empirically evaluating human–ai work systems. *AI Magazine*, 44(3):274–281. DOI: https://doi.org/10.1002/aaai.12108.

Leben, D. (2025). *AI Fairness: Designing Equal Opportunity Algorithms*. The MIT Press.

Marangon, J. D. (2025). Google cloud vision api for image handling and ocr. https://medium.com/@johnidouglasmarangon/google-cloud-vision-api-for-image-handling-and-ocr-a6763969a2e6, Accessed: 10 July 2026.

Martinez-Martin, N. (2019). What are important ethical implications of using facial recognition technology in health care? *AMA Journal of Ethics*, 21(2):E180–187. DOI: https://doi.org/10.1001/amajethics.2019.180.

McLymore, A. (2025). Amazon cuts reference to diversity from annual report. https://www.reuters.com/technology/amazon-cuts-reference-diversity-annual-report-2025-02-07/, Accessed: 10 July 2026.

Mohammadi, E., Cai, Y., Novin, A., Vera, V., and Soltanmohammadi, E. (2025). Who is a scientist? gender and racial biases in google vision ai. *AI and Ethics*, 5(5):4993–5010. DOI: https://doi.org/10.1007/s43681-025-00742-4.

Moon, D. and Ahn, S. (2025). Metrics and algorithms for identifying and mitigating bias in ai design: A counterfactual fairness approach. *IEEE Access*, 13:59118–59129. DOI: https://doi.org/10.1109/ACCESS.2025.3556082.

Nunes, J. L., Barbosa, G. D. J., de Souza, C. S., and Barbosa, S. D. J. (2024). Using model cards for ethical reflection on machine learning models: an interview-based study. *Journal on Interactive Systems*, 15(1):1–19. DOI: https://doi.org/10.5753/jis.2024.3444.

Oyěwùmí, O. (2021). *A invenção das mulheres: Construindo um sentido africano para os discursos ocidentais de gênero*. Bazar do Tempo. https://www.bazardotempo.com.br/products/a-invencao-das-mulheres, Accessed: 14 July 2026.

Paganoni, M. C. *et al*. (2019). Ethical concerns over facial recognition technology. *Anglistica AION an Interdisciplinary Journal*, 1:83–92. DOI: https://doi.org/10.19231/angl-aion.201915.

Parks, M. (2025). The ethics of ai at the intersection of transgender identity and neurodivergence. *Discover Artificial Intelligence*, 5(1):34. DOI: https://doi.org/10.1007/s44163-025-00257-1.

Perilo, M., Valença, G., and Telles, A. (2024). Non-binary and trans-inclusive ai: A catalogue of best practices for developing automatic gender recognition solutions. *SIGAPP Appl. Comput. Rev.*, 24(2):55–70. DOI: https://doi.org/10.1145/3687251.3687255.

Quaresmini, C. and Zanotti, G. (2025). Fairness through feedback: Addressing algorithmic misgendering in automatic gender recognition. *arXiv*. DOI: https://arxiv.org/abs/2506.02017.

Quinan, C. and Hunt, M. (2022). Biometric bordering and automatic gender recognition: Challenging binary gender norms in everyday biometric technologies. *Communication, Culture and Critique*, 15:211–226. DOI: https://doi.org/10.1093/ccc/tcac013.

Rafael, R. d. M. R., Santos, H. G. d. S., Caravaca-Morera, J. A., Wilson, E. C., and Breda, K. L. (2023). Inclusão ou ilusão da identidade de gênero no país com o maior número de assassinatos de transgêneros: um ensaio crítico brasileiro. *Escola Anna Nery*, 27:e20230117. DOI: https://doi.org/10.1590/2177-9465-EAN-2023-0117pt.

Raza, S., Shaban-Nejad, A., Dolatabadi, E., and Mamiya, H. (2024). Exploring bias and prediction metrics to characterise the fairness of machine learning for equity-centered public health decision-making: A narrative review. *IEEE Access*, 12:180815–180829. DOI: https://doi.org/10.1109/ACCESS.2024.3509353.

Reis, V. Q. d., Rabello, M. E. R., Lima, A. C., Jardim, G. P. S., Fernandes, E. R., and Brefeld, U. (2023). Data practices in apps from brazil: What do privacy policies inform us about? *Journal on Interactive Systems*, 14(1):1–8. DOI: https://doi.org/10.5753/jis.2023.2954.

Reuters (2025). Amazon ending dei programs. `https://www.theguardian.com/technology/2025/jan/10/amazon-ending-dei-programs`, Accessed: 10 July 2026.

Santos, L. G. d. M., Costa, A. B. d., David, J. d. S., and Pedro, R. M. L. R. (2023). Reconhecimento facial: Tecnologia, racismo e construção de mundos possíveis. *Psicologia Sociedade*, 35:e277141. DOI: https://doi.org/10.1590/1807-0310/2023v35e277141.

Scheuerman, M. K., Paul, J. M., and Brubaker, J. R. (2019). How computers see gender: An evaluation of gender classification in commercial facial analysis services. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–22. Article 144. DOI: https://doi.org/10.1145/3359643.

ScienceDaily (2025). Google's deepfake hunter sees what you can't - even in videos without faces. `https://www.sciencedaily.com/releases/2025/07/250724232412.htm`, Accessed: 10 July 2026.

Scott, J. W. (1986). Gender: A useful category of historical analysis. *The American Historical Review*, 91(5):1053–1075. DOI: https://doi.org/10.2307/1864376.

Silva, H. H. (2021). Algoritmos de reconhecimento facial e as discriminações contra pessoas transexuais. *Internet & Sociedade*, 2(2):47–66. https://revista.internetlab.org.br/wp-content/uploads/2022/03/Algoritmos-de-reconhecimento-facial-e-as-discriminacoes-contra-pessoas-

transexuais.pdf, Accessed: 10 July 2026.

Silva, L. K. M. d., Silva, A. L. M. A. d., Coelho, A. A., and Martiniano, C. S. (2017). Uso do nome social no sistema Único de saúde: elementos para o debate sobre a assistência prestada a travestis e transexuais. *Physis: Revista de Saúde Coletiva*, 27(3):835–846. Accessed: 10 July 2026. DOI: https://doi.org/10.1590/S0103-73312017000300023.

Silva, T. (2022). *Racismo algorítmico: inteligência artificial e discriminação nas redes digitais*. Edições Sesc SP. https://racismo-algoritmico.pubpub.org, Accessed: 10 July 2026.

Soldatelli, G. D. (2022). Uma avaliação sobre o viés de gênero em reconhecimento facial. Master's thesis, Universidade Federal de Mato Grosso. http://bdm.ufmt.br/handle/1/3949, Accessed: 10 July 2026.

Spizzirri, G., Eufrásio, R., Lima, M. C. P., de Carvalho Nunes, H. R., Kreukels, B. P. C., Steensma, T. D., and Abdo, C. H. N. (2021). Proportion of people identified as transgender and non-binary gender in brazil. *Scientific Reports*, 11(1):2240. DOI: https://doi.org/10.1038/s41598-021-81411-4.

Tafesse, E. (2025). Tech companies are abandoning diversity to satisfy trump. it will hurt them in the long run. https://www.techpolicy.press/tech-companies-are-abandoning-diversity-to-satisfy-trump-it-will-hurt-them-in-the-long-run/, Accessed: 10 July 2026.

Turchi, T., Malizia, A., and Borsci, S. (2024). Reflecting on algorithmic bias with design fiction: The minicode workshops. *IEEE Intelligent Systems*, 39(02):40–50. DOI: https://doi.org/10.1109/MIS.2024.3352977.

Union, A. C. L. (2018). Amazon's face recognition falsely matched 28 members of congress with mugshots. https://www.aclu.org/news/privacy-technology/amazons-face-recognition-falsely-matched-28, Accessed: 10 July 2026.