Gender in Pixels: Pathways to Non-binary Representation in Computer Vision

Elena Beretta
Vrije Universiteit Amsterdam
De Boelelaan 1105, 1081 HV
Amsterdam, Netherlands
elena.beretta@vu.nl

Abstract

In the field of Computer Vision (CV), the study of bias, including gender bias, has received a significant area of attention in recent years. However, these studies predominantly operate within a binary, cisnormative framework, often neglecting the complexities of non-binary gender identities. To date, there is no comprehensive analysis of how CV is addressing the mitigation of bias for non-binary individuals or how it seeks solutions that transcend a binary view of gender. This systematic scoping review aims to fill this gap by analyzing over 60 papers that delve into gender biases in CV, with a particular emphasis on non-binary perspectives. Our findings indicate that despite the increasing recognition of gender as a multifaceted and complex construct, practical applications of this understanding in CV remain limited and fragmented. The review critically examines the foundational research critiquing the binarism in CV and explores emerging approaches that challenge and move beyond this limited perspective. We highlight innovative solutions, including algorithmic adaptations and the creation of more inclusive and diverse datasets. Furthermore, the study emphasizes the importance of integrating gender theory into CV practices to develop more accurate and representative models. Our recommendations advocate for interdisciplinary collaboration, particularly with Gender Studies, to foster a more nuanced understanding of gender in CV. This study serves as a pivotal step towards redefining gender representation in CV, encouraging researchers and practitioners to embrace and incorporate a broader spectrum of gender identities in their work.

Introduction

As the capabilities of Artificial Intelligence (AI) expand, our society faces increasingly complex challenges and opportunities at the intersection of technology and human identity. This evolution of AI brings to the forefront questions about how technology recognizes, interprets, and responds to the diverse human identity traits (Scheuerman et al. 2020; Monea 2022). In particular, the interaction between AI systems and the multifaceted nature of human identity becomes a crucial area for ethical and social examination. In the domain of Computer Vision (CV), this interaction is acutely pronounced (Schwemmer et al. 2020). CV systems, designed to interpret and understand visual data from the physical world, are increasingly entwined with human identity recognition (Díaz et al. 2023). These systems have found daily applications in a diverse array of fields, ranging from demographic analysis (Lin, Kim, and Joo 2022) and commercial applications (Ovalle, Liang, and Boyd 2023) to interactive media (Akhyani et al. 2022) and surveillance (Scheuerman, Pape, and Hanna 2021). However, this rapid growth and integration of CV in into multiple sectors not only brings forth critical ethical considerations (Cabello et al. 2023), but also highlights a growing awareness of these systems' potential to exclude or misrepresent minorities and marginalized communities, as evidenced by the increasing focus on studies addressing algorithmic fairness and social bias in CV (Collett and Dillon 2019; Heldreth et al. 2023; Mousavi, Shahbazi, and Asudeh 2024; Quaresmini and Primiero 2023). This concern is particularly relevant in the context of gender bias, where traditional CV models have predominantly operated within a binary and cisnormative gender framework (Scheuerman and Brubaker 2018). Such a simplified categorization often overlooks the rich spectrum of gender identities, leading to the exclusion and misrepresentation of non-binary, gender nonconforming, and transgender individuals (Barlas et al. 2021; Hoefsloot 2021). This not only raises concerns about the social and psychological impacts of these technologies but also prompts a reevaluation of how these systems are designed and deployed (Scheuerman, Paul, and Brubaker 2019). The limitations of binary gender classification in CV are not merely technical oversights but reflect deeper cultural and societal narratives embedded within these technologies (Cairns 2021). For instance, the reliance on physical markers for gender predictions overlooks the performative and non-fixed nature of gender identity. As gender identity is a result of a complex interplay of biological, psychological, and social factors, its representation in machine vision demands a nuanced understanding and approach (Keyes 2018). Recognizing these challenges, our investigation contributes to the broader academic dialogue on the societal impacts of CV technology, focusing on the complexities of gender identity. Our research aims to conduct a systematic scoping review on encompassing non-binary gender perspectives in CV. This review explores not only the prevalence of gender biases but also examines the kind of solutions, critical studies, and inclusive approaches that have been proposed to foster inclusivity in this field. By delving into various academic papers and empirical studies, we aim to understand how the field of CV is evolving to accommodate a more inclusive understanding of gender and what methodologies are being adopted to mitigate these biases. Our investigation contributes to the broader academic dialogue on the societal impacts of CV technology, focusing on the complexities of gender identity. By addressing the implications of binary gender perspective in CV, we hope to highlight the need for systems that are technically proficient while being deeply attuned to the intricacies of human identity. This research underscores the imperative for a holistic approach in the development and deployment of CV technologies, one that extends beyond technical solutions to encompass ethical and societal considerations.

Background

Related Work

Galai's work (Galai 2023) provides a critical lens on the socio-political dimensions influencing visual technologies, including CV. It does not specifically focus on bias mitigation or fairness in CV but rather offers a critique of the embedded biases in these technologies, particularly around race and gender. The paper underscores the historical and systemic biases, such as White-centrism and transphobic visualities, that shape visual technologies. While Galai does not directly conceptualize or operationalize gender in CV, the analysis serves to understand how societal prejudices can be reflected in CV technologies. Fabbrizzi et al. (Fabbrizzi et al. 2022) provides an in-depth examination of biases in CV. It emphasizes the significance of addressing biases to ensure fairness and ethical considerations in CV. The paper is not solely focused on gender but addresses a broader range of identity traits and sensitive attributes. It explores the concept of gender and other protected attributes in the context of CV, discussing how these are operationalized and represented in visual datasets. The paper also reviews various approaches to bias mitigation, including the development of bias-aware data collection methods and the proposal of a checklist to identify different types of biases during dataset collection. This work contributes to understanding and addressing biases in CV, particularly concerning the representation of diverse human identities.

The work by Lee et al. (Lee et al. 2023) extends the dialogue on biases in CV systems by examining the representation of non-binary genders. Their study delves into how CV technologies often neglect or inaccurately represent non-binary identities, primarily due to reliance on binary gender models. This research underscores the need for CV systems to adopt more inclusive approaches that accurately reflect the diversity of gender expressions.

Udefi et al. (Udefi et al. 2023) offer a comprehensive examination of biases present in facial image processing systems, particularly focusing on the representation of various races, ethnicities, cultures, ages, and genders. The study investigates the origins of these biases, whether stemming from the algorithms themselves or from the datasets used. A critical aspect of their analysis is the exploration of the underrepresentation of certain groups in widely used datasets. This work contributes to the understanding of dataset biases in computer vision, although not directly focused on gender beyond binarism.

Integrating these perspectives with (Leslie 2020), we see a contribution to this discourse by highlighting the ethical considerations in the design and deployment of CV technologies. The work emphasizes the impact of systemic biases in visual technologies and discusses the potential consequences of these biases on marginalized groups. This work calls for a more responsible and inclusive approach in CV, stressing the importance of considering the broader societal implications of these technologies.

Gender as Performative Act

In the prevailing Western cultural framework, the concepts of sex and gender are often intertwined yet distinguished in unclear ways (Scheuerman, Pape, and Hanna 2021). Sex refers to a biological classification (male or female), typically based on physical attributes and genetic makeup. Gender, on the other hand, is understood as a societal role or identity (man or woman) shaped by cultural norms and individual behavior. This conventional perspective implies that gender is a direct outcome of one's biological sex. Such a view leads to the perception of gender as dualistic (limited to men and women), unchangeable (fixed at birth and unalterable), and rooted in physical traits (determined and expressed through bodily characteristics). This traditional understanding influences societal interactions and perceptions of gender identity. While this binary interpretation is widely recognized as a simplistic and inaccurate representation, it continues to dominate the approach to gender acknowledgement in the field of CV.

As proposed by Judith Butler (Butler 1988, 2006), gender performativity suggests that gender identity is an ongoing, interactive process shaped by societal norms, cultural contexts, and individual actions. In CV, this concept necessitates a shift from static, binary gender classifications to more fluid and context-sensitive interpretations (Weinschenk 2021). The challenge lies in designing CV systems that can accommodate the fluidity and performative nature of gender. This involves recognizing gender as a spectrum rather than a binary and understanding that gender expression may not always align with traditional physical markers (Rosenbaum 2022). It also requires acknowledging the impact of social and cultural contexts on how gender is expressed and perceived.

Methodology

Scope of the Study

This systematic scoping review is dedicated to conducting an in-depth exploration of perspectives that extend beyond binary gender models and cisnormative assumptions in CV. The objective of our research is to scrutinize how CV technologies and methodologies currently engage with a broad spectrum of gender identities, and to assess the extent to which non-binary gender perspectives are acknowledged and incorporated in these technologies and applications. The

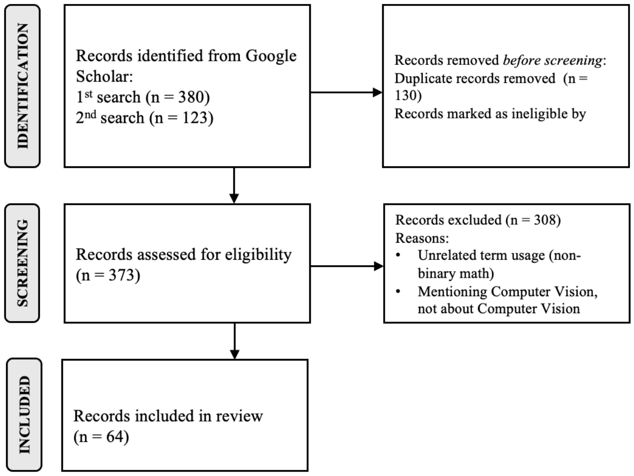

Figure 1: PRISMA Informed Process Flow demonstrating how we searched for, screened, and selected for inclusion the papers in our review

review delves into a variety of studies and literature, seeking to uncover not just the challenges posed by these biases, but also the solutions, critical analyses, and inclusive methodologies proposed to counteract them and foster a more inclusive environment within the field of CV. Figure 1 shows a PRISMA diagram (Page et al. 2021), which outlines the flow of our search and selection process.

Identification

The beginning of our review was characterized by a focused preliminary analysis, aimed at understanding the landscape of gender terminology within the fields of CV. This initial phase involved examining a subset of papers that explicitly mentioned gender-related terms in their titles. The purpose of this was twofold: to measure the level of specificity with which gender terms are employed within the existing body of research and to identify the most relevant and frequently used terms that could guide our comprehensive search strategy. Our findings from this preliminary review indicated a notable trend: the discourse within CV research often employs gender terms in a broad or generic manner, with limited specificity. Terms such as ”non-binary,” ”transgender,” and similar broad descriptors were commonly used, whereas more nuanced or specific gender identities were less frequently articulated. This observation was instrumental in shaping our search approach. It led us to the conclusion that incorporating wide-ranging and inclusive gender terms as keywords would be the most effective strategy to capture the breadth of research concerning gender inclusivity and bias in this domain. Consequently, our search was structured around these broad terms, reflecting the generalized usage of gender terminology within the field. Given this context, our search through Google Scholar was executed with a set of keywords derived from the initial review. This strategy was designed to ensure inclusivity and comprehensiveness, capturing studies that, while perhaps not detailed in their definition of gender, contribute significantly to discussions on gender bias, inclusivity, and diversity within CV. The search unfolded in two rounds:

- **Initial Search (August 2023)**: resulted in 380 records, providing a foundational overview of relevant literature;
- **Follow-up Search (December 2023)**: brought in an additional 123 records, enriching our dataset with the most recent scholarly contributions.

Altogether, 503 records were identified, laying the groundwork for our systematic review. To maintain transparency and facilitate replication of our search methodology, the keywords employed are documented in Table 1.

Data Screening

Following the identification process which garnered 503 records, the initial step of our data screening was the removal of duplicate entries, which resulted in the exclusion of 130 records. This action reduced the pool to 373 records. As we progressed with the screening, a focused evaluation led to the exclusion of 308 records. The exclusion was due to two main reasons:

- Unrelated Term Usage: a portion of the excluded records employed terms such as ”non-binary” in contexts unrelated to gender identity, such as mathematical or computational descriptions (”non-binary math”);
- Peripheral Mention of Computer Vision: several records were excluded because, although they mentioned ”computer vision” within the text, their primary focus was not on computer vision technology or its applications in the context of gender bias and inclusivity. This included studies where computer vision was cited as a tangential reference rather than the core subject of investigation.

The culmination of this rigorous screening process was the selection of 65 papers deemed to directly align with our research objectives. These papers collectively offer valuable insights into the prevailing challenges, innovative methodologies, and solutions geared towards fostering gender inclusivity in CV.

Data Organization

In constructing the framework for our review, we established a priori categories for organizing the selected studies. The decision was driven by the intent to create a coherent structure that could accommodate the breadth and depth of research in gender inclusivity within CV. Importantly, this categorization does not imply mutual exclusivity among the selected studies.

- **Foundational research**: the review encompasses papers that offer a critical examination of both binary and cisnormative models of gender in CV, specifically those with a gender focus;
- **Bias Mitigation and fostering inclusive solutions beyond binarism and cisnormativity**: we included studies that address bias mitigation and foster inclusivity, acknowledging non-binary genders and other sensitive attributes. This category also covers research that, while not exclusively focused on gender, contributes to broader inclusivity beyond binary and cisnormative paradigms;

Generic terms	Methods specific	Gender-specific	Technology specific
image	debiasing	non-binary	face processing
computer vision	bias mitigation	transgender	face analysis
gender	fairness	gender nonconforming	face detection
bias			face recognition

Table 1: Terms used to search Google Scholar

- **Applications beyond binarism**: We also considered works that propose CV applications transcending binary models. These include studies that may not directly tackle gender bias but offer indirect solutions or are inclusive of a broad non-binary spectrum without necessarily problematizing gender issues.

Inclusion and Exclusion Criteria

The criteria for selecting articles were as follows (prior to data identifiction):

- Articles must explicitly acknowledge non-binary gender identities (or a broader spectrum), whether the main focus is on gender bias or not.
- We focused on articles that specifically discuss bias mitigation in CV with a non-binary gender perspective.
- To maintain the review's focus, we excluded general articles on bias mitigation in CV that did not specifically address gender (among other sensitive attributes) and dit not acknowledge non-binary and/or cisnomative perspectives.

This systematic approach ensures that our review comprehensively covers the landscape of non-binary gender perspectives in CV, from foundational critiques of binary and cisnormative models to innovative applications that broaden the scope of inclusivity.

Results

General Results

Type of Contribution The nature of contributions offered by the papers discussed in Section has been summarized in Table 2. The majority of contributions fall into the category of *bias mitigation and fostering inclusive solutions* (see Table 2. This category, comprising nearly half of the total papers, may suggest a growing commitment to addressing gender bias and advancing inclusivity beyond binarism within the field.

Gender Acknowledgement Graph 2 illustrates the progressive increase in the number of research papers published over the years that acknowledge gender diversity beyond the binary paradigm in the field of CV (see also Table 3. It is important to note that the articles considered encompass a range of acknowledgments but may not necessarily propose specific solutions or represent critical studies; their significance lies in their recognition of diverse gender identities. Over the years, the field has seen a remarkable increase in the proportion of such articles, signifying a growing awareness of gender diversity within the CV research community.

Gender Acknowledgement Evolution in Computer Vision Research

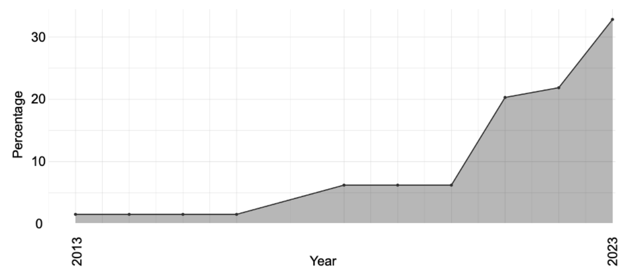

Figure 2: Chart showing the growth of gender-diverse acknowledgments in computer vision research over time

Main Contributions Table 4 categorizes and provides a concise overview of the key contributions within the surveyed papers.
Data: This category encompasses contributions that primarily focus on data-related solutions. These papers aim to address gender diversity concerns by tackling issues related to data collection, processing, or representation.
Algorithm: The category includes contributions that extend beyond binary gender models through algorithmic solutions.
Reasoning/Theory: Papers falling under this category contribute new theories, critical perspectives, and advanced reasoning beyond binary gender paradigms.
Impact/Audit: This category comprises contributions that focus on impact measurement or audit analysis in the context of gender diversity.

Type of Technology and Applications Table 5 explores the diverse range of technologies and their practical applications as identified in the surveyed papers. There is a foundational presence of Generic AI and Computer Vision studies. Face technology shows a varied application with Face Recognition and Face Analysis/Attribute Classification each constituting 9.38% of the focus. Annotation technologies, including Image Tagging and Vision-Language Models, account for a significant part, each marking around 4.69%.

Inclusive Practices

This section delves into the inclusive practices identified in the selected papers from our survey, as summarized in Table 6. These practices represent efforts within the field of CV to

Type of Contribution	Inclusion Criteria	Amount	Percentage
Foundational research	Critical studies	16	25%
Bias mitigation	Inclusive solutions in CV (beyond binarism)	31	48.4%
Applications	Considers gender beyond binarism in CV applications	12	18.75%
Gender focused	Considers gender beyond binarism in CV	20	31.5%

Table 2: The general inclusion method used in the surveyed papers. It is important to acknowledge that some papers may fall under multiple categories, resulting in counts that do not equate to the overall total.

Gender Acknowledgement	Inclusion Criteria	Amount	Percentage
Neutral	Embraces the concept of ”neutrality” (mixed groups or unknowns)	2	3.13%
Non binary	Considers beyond the two genders	50	78.12%
Transgender	Considers beyond cisnormativity	27	42.19%
Broad spectrum	Considers a broader spectrum beyond the two genders and cisnormativity	12	18.75%

Table 3: The gender processing method used in the surveyed papers. It is important to acknowledge that some papers may fall under multiple categories, resulting in counts that do not equate to the overall total.

extend beyond traditional binary gender models and cisnormative assumptions, aiming to create a more inclusive and representative technological environment. Notably, the selection includes all papers where the authors have declared their goal to promote inclusivity in CV.

Beyond Binary Classification Various methodologies have been explored to predict the gender of non binary and/or transgender individuals. These approaches vary in their focus and the aspects of facial recognition they highlight. One approach involves analyzing facial features in relation to appearance and therapy factors for transgender individuals (Kumar et al. 2016; Vijayan, Kareem, and Kizhakkethottam 2016). These studies examine how changes in appearance, possibly due to therapy, affect the accuracy of gender recognition algorithms. Another branch of the field investigates the reliability of different facial regions for gender prediction. In particular, Mahalingam et al. (Mahalingam and Ricanek 2013) specifically compare the effectiveness of using the eye region against full-face recognition in a transgender dataset. In a subsequent study (Mahalingam, Ricanek, and Albert 2014), they focus on the periocular region of the face, assessing the accuracy of this specific area in recognizing gender, especially when individuals undergo gender transformation. Alongside these specific approaches, there are also efforts more broadly oriented towards extending gender recognition beyond binary classifications. In (Wu et al. 2020), the authors propose a method to refine gender classification while simultaneously mitigating inherent biases within facial image datasets. This approach includes a detailed analysis and application of algorithmic adjustments to these datasets, aiming to enhance the representation of gender in CV. In (Chin-Purcell and Chambers 2021) the authors focus on the use of CNNs to address disparities in gender classification. This research involves an extensive process of training and testing CNN models on a variety of datasets, identifying and addressing any gaps in accuracy, especially in terms of gender classification.

Blindness Approach In contrast to the previous practice, this approach involves the deliberate omission of gender as a classification feature in algorithmic processes; it stems from the assumption that eliminating gender from data analysis can potentially reduce the perpetuation of biases. In (Herruzo 2021), the discussion centers around the ethical implications of surveillance and the importance of maintaining privacy, especially in public spaces. The paper argues for a CV approach that respects individual privacy, which includes not identifying or classifying individuals based on gender, thereby reducing bias and upholding ethical standards. Similarly, in (Ravfogel et al. 2022) the authors propose a practice where gender information is intentionally ’erased’ from data used in machine learning models. This method aims to prevent the model from learning and perpetuating gender-based biases. By removing gender as a variable, the study investigates the potential for creating more neutral and unbiased algorithms in CV.

Self-reported Gender The approach of self-reported gender is gaining traction as a method that respects and acknowledges the individual’s perspective on their gender identity. Studies focusing on dataset creation (Hazirbas et al. 2022) and bias measurement (Porgali et al. 2023) advocate for consent-driven and ethical data collection practices, highlighting how self-reported gender can contribute to creating more representative and equitable CV models. In (Thong, Joniak, and Xiang 2023; Schumann et al. 2023), the authors explore the multidimensionality of skin tone and its relationship with gender perception. In (Andrews et al. 2023b; Lee et al. 2023) the authors further discuss the principles guiding ethical data curation and the presence of social biases in vision-language models. Both highlight the significance of self-reported gender in mitigating biases and maintaining ethical standards in CV. Moreover, (Keyes 2018) and (Ranjit et al. 2023) address the challenges of automatic gender recognition systems and the impact of model fine-tuning on gender biases. These studies point to the potential of self-

Main Contribution	Criteria	Amount	Percentage
Data	Contribution that focuses on data-related solutions	16	25%
Algorithm	Algorithmic solutions beyond binarism	27	42.2%
Reasoning/theory	New theories, critiques, and reasoning beyond binarism	14	21.3%
Impact/Audit	Impact measurement or audit analysis	20	31.5%

Table 4: Contributions provided by surveyed papers. It is important to acknowledge that some papers may fall under multiple categories, resulting in counts that do not equate to the overall total.

reported gender data in reducing misgendering errors and adjusting biases in visual recognition models. Lastly, (Katzman et al. 2023) examines the adverse effects of misrepresentation in image tagging, suggesting that self-reported gender can play a critical role in minimizing these harms.

Informed Consent Informed consent involves obtaining explicit permission from individuals before their data, especially sensitive information, is used for algorithmic processing. It ensures that users and/or participants are aware of how their data is used and grants them control over their personal information. Several studies have highlighted the importance of this practice in ensuring ethical standards and respecting the autonomy of individuals. Both (Hazirbas et al. 2022) and (Porgali et al. 2023) emphasize the need for consent-driven data collection methods. They showcase the design and implementation of large datasets specifically created with the consent of participants, ensuring that the data used to measure algorithmic bias and robustness is ethically sourced and consented. Similarly, (Andrews et al. 2023a) discusses the ethical challenges and considerations involved in collecting image datasets that feature human subjects. The paper argues for the necessity of informed consent, especially when dealing with human-centric images. Additionally, (Schumann et al. 2023) delves into the subjectivity of skin tone annotation in machine learning. This study underlines the importance of informed consent in collecting and annotating data related to skin tone and gender.

Gender as Discrete Variable The approach to gender in CV relies on dichotomous, binary assumptions. However, some recent studies are moving towards a more nuanced understanding, considering gender as a discrete, non-binary variable. This shift acknowledges that gender, often perceived and classified from an external perspective, can benefit from a more continuum-based approach. In this context, gender is labeled as *perceived gender presentation* and is operationalized by visual traits that stereotypically belong to the male or female gender, essentially based on the visual traits of perceived gender. This perspective aligns with the discussions in (Kirchler et al. 2021) about adapting interactive systems in explainable AI for gender recognition. In the realm of bias mitigation and audit, methods like those proposed in (Georgopoulos et al. 2021) for reducing bias in facial datasets involve transferring attributes across various demographics. Similarly, (Gustafson et al. 2023) introduces a benchmarking tool designed to evaluate fairness in CV systems.

Participatory Design Participatory design within the domain of CV, particularly in the context of Automatic Gender Recognition (AGR), presents an innovative approach to incorporating diverse and complex identities into algorithm development. This methodology involves engaging both the designers of AGR algorithms and the potential users, especially those from marginalized groups, in the design process. The goal is to bridge the gap between algorithm designers and marginalized users, and to develop methods that better conceptualize gender identity within algorithms. A significant contribution in this area is represented by (Scheuerman and Brubaker 2018). The study is centered on the design and impacts of AGR on transgender people, launching participatory design workshops to involve both the creators of AGR algorithms and their potential users. The objective is to create a dialogue between the needs and concerns (Hamidi, Scheuerman, and Branham 2018) of the *targets* of these systems and the constraints and work practices of the designers, which is a crucial step towards achieving more inclusive and representative algorithmic solutions, but also to empower users with customizability and fluidity, including the option to opt out of categorization entirely.

Reasons for Exclusion

The reasons for exclusion of non-binary identities are diverse, each pointing to different aspects of research design and data handling (see Table 7). Firstly, a prevalent reason is the lack of data (Booth et al. 2021; Meister et al. 2023; Hoque et al. 2020; Keswani and Celis 2021; Dominguez-Catena, Paternain, and Galar 2023; Hirota, Nakashima, and Garcia 2022; Alasadi, Al Hilli, and Singh 2019). Numerous studies point to an absence or scarcity of data representing non-binary individuals. Another significant reason is statistical concerns (Booth et al. 2021; Lin, Kim, and Joo 2022; Qi et al. 2022). Several studies have excluded non-binary participants due to insufficient numbers, making statistical analysis challenging and potentially unreliable. This issue often arises in studies focused on bias or fairness, where the limited sample size of non-binary individuals leads to their exclusion. In cases where studies are auditing existing models or datasets that do not include non-binary data, the inclusion of non-binary identities is considered impractical (*Unfeasibility caused by the task*) (Wang et al. 2022, 2019; Smith et al. 2023; Shen et al. 2023; Meister et al. 2023; WANG et al. 2023; Mandal, Little, and Leavy 2023; Hoque et al. 2020; Ranjit et al. 2023; Alasadi et al. 2022; Schumann et al. 2021; Dominguez-Catena, Paternain, and Galar 2023). This limitation often stems from the nature of the task itself and the existing frameworks within which these studies operate. Furthermore, some studies explicitly recognize the exclusion of non-binary identities as a study limitation (Wang

Type of Technology and Applications	Amount	Percentage
General		
Generic AI	1	1.56%
Generic CV	10	15.63%
Data		
Unstructured Data	1	1.56%
Visual Dataset	4	6.25%
Facial Dataset	1	1.56%
Face Technology		
Face Recognition	6	9.38%
Face Verification/Matching	2	3.13%
Face Analysis/Attribute Classification	6	9.38%
Emotions and Facial Expression Recognition	2	3.13%
AGR	7	10.94%
Annotation		
Image tagging	3	4.69%
Image Captioning	1	1.56%
Video Annotation	1	1.56%
Vision-Language Models	3	4.69%
Task		
Explainable AI/CV	1	1.56%
GAN	1	1.56%
Image Search	1	1.56%
Knowledge Extraction	1	1.56%
Machine Learning	3	4.69%
Application		
Cyberbullying	1	1.56%
Hiring	1	1.56%
Security and Surveillance	2	3.13%
Commercial Gender Classification and Face Analysis Systems	4	6.25%
Art Work	2	3.13%
Makeup Support System	1	1.56%
Mobile Biometric	1	1.56%
Live Telecast	1	1.56%

Table 5: Type of Technology and Applications provided by surveyed papers. It is important to acknowledge that some papers may fall under multiple categories, resulting in counts that do not equate to the overall total.

Inclusive Practices	Amount	Percentage
Beyond Binary Classification	6	9.4%
Blindness Approach	2	3.12%
Self-reported gender	10	15.62%
Informed Consent	4	6.25%
Gender as Discrete Variable	3	4.68%
Participatory Design	1	1.56%

Table 6: Overview of the inclusive practices proposed in the surveyed papers. It is important to acknowledge that some papers may fall under multiple categories, resulting in counts that do not equate to the overall total.

et al. 2022, 2019; Smith et al. 2023; Mandal, Little, and Leavy 2023; Hoque et al. 2020; Alasadi et al. 2022). These acknowledgments indicate an awareness of the need for broader inclusivity but also highlight the existing challenges and constraints within research methodologies and available data. Looking towards the future, many studies point to the inclusion of non-binary identities as an objective in future work or recommendations (Alasadi et al. 2020; Mandal, Little, and Leavy 2023; Akhyani et al. 2022; Muthukumar, Kingsbury, and Mojsilović 2018). Finally, there are cases where studies are observed without motivation for exclusion (Villalobos, Mery, and Bowyer 2022; Teotia, Lapedriza, and Ostadabbas 2022; Ravfogel et al. 2022; Muthukumar et al. 2019; Akhyani et al. 2022; Muthukumar, Kingsbury, and Mojsilović 2018). In these cases, while the existence of non-binary identities is acknowledged, no explicit reasoning is provided for their exclusion.

Reasons for Exclusion	Amount	Percentage
Lack of data	7	10.94%
Statistical Reasons	3	4.68%
Unfeasibility caused by the Task	12	18.75%
Study Limitation	6	9.37%
Future Work/Recommendations	4	6.25
Without Motivation	6	9.37%

Table 7: Overview of the reasons why non-binary identities were not considered in the studies and/or were excluded in the surveyed papers. It is important to acknowledge that some papers may fall under multiple categories, resulting in counts that do not equate to the overall total.

Discussion

Gender Data Annotation: Inclusivity and Perception

There is an increasing emphasis on using self-reported data as a foundation for inclusivity. The commitment to representing gender data through annotation strives for inclusivity but encounters notable challenges. Self-reported annotations are invaluable, providing a direct line to how individuals identify themselves across a spectrum of identities, including non-binary. However, as (Schumann et al. 2023) indicates, these data sets are not without their drawbacks. They tend to be costly and often too *clean*, lacking the complexity found in everyday, *in the wild* contexts. This distinction is crucial in understanding the interaction between models and the identities of users versus the performance across perceived attributes in diverse imagery, including generated images.

Moving from the individual to the collective (Leslie et al. 2022) warns of the risks inherent in data design—especially when categories of identity are overly simplified or grouped together. This simplification can result in the erasure of distinctive identity claims for those who fall outside the majority, such as non-binary and trans individuals, or those from diverse racial backgrounds who find themselves subsumed under broad and nondescript categories like *non-white*.

The delicate act of balancing visibility with accuracy is further complicated by privacy concerns. (Xiang 2022) addresses the tension between being seen for fairness and the risk of being mis-seen, leading to privacy invasions and unfair treatment. The collection of diverse data must be balanced with informed consent to protect privacy while mitigating biases that may arise from misidentification or misrepresentation.

Gender in Visual Data

The landscape of bias within CV is marked by its complexity, a duality ingrained in the *first level of representation* (Schwemmer et al. 2020). At this fundamental level, we ask: does the algorithm recognize individuals across gender with uniform accuracy? (Noble 2018) spotlights this concern as a foundation of the algorithmic bias discourse, framing injustice and discrimination in terms of error rates that disproportionately impact certain social groups. Still, bias transcends identification accuracy and delves into the content algorithms produce, such as the systematic variation in label subsets corresponding to different demographics (Corbett-Davies et al. 2017). However, equal recognition is only the surface. The second dimension of bias, often overlooked, is content bias (Schwemmer et al. 2020). This bias becomes apparent when algorithms output biased label subsets that fail to accurately reflect the diversity and the complexity of visual data within demographic groups.

In response to these biases, the field has explored *fairness-through-blindness* approaches, which focus on removing visual artifacts associated with gender (Keyes 2018; Ravfogel et al. 2022). Yet, as (Meister et al. 2023) underscores, the attempts to remove bias by erasing gender cues from the person's appearance approaches often fall short, as gender artifacts are deeply embedded in the content and context of images visual datasets. By attempting to strip images of gendered elements, we inadvertently endorse the very stereotypes we seek to eliminate and ignore the fact that these elements are intertwined with other informative aspects of the image. Such practices can perpetuate biases and overlook the complexity of gender as a social construct that cannot be distilled into simple visual artifacts (Scheuerman, Pape, and Hanna 2021).

Recognizing the complexity of gender representation leads us to favor the notion of *perceived gender* (Kirchler et al. 2021; Gustafson et al. 2023). This perspective considers that gender, as a construct, cannot be accurately perceived or annotated based on visual data alone. It respects the subjective nature of gender identity and challenges the reliance on visual cues, which are prone to reinforce societal stereotypes (Wang et al. 2022; Scheuerman, Pape, and Hanna 2021).

Despite these insights, some authors argue against the possibility of completely debiasing vision-language data due to the absence of unbiased gender data. The current state of the art is constrained by the data it has to work with, which is inherently biased (Smith et al. 2023). This limitation brings to light the importance of developing innovative methods like synthetic contrast sets to balance representations within the data, despite acknowledging that a perfect solution may not be attainable.

Anticlassification

The complexity of bias within CV underscores the inadequacy of conventional gender prediction practices. The pursuit of uniform algorithmic recognition, while pivotal, does not address the full scope of bias. It becomes clear that solutions cannot be found in the same methods that confine gender to binary visual artifacts, which are deeply embedded in the content and context of images. This acknowledgment paves the way for an anticlassification approach that opposes the practice of gender prediction (Thong, Joniak, and Xiang 2023; Booth et al. 2021; Meister et al. 2023; Scheuerman and Brubaker 2018; Hamidi, Scheuerman, and Branham 2018), especially those extending beyond binary classifications that claim, but fail, to promote inclusivity. Aligning with the insights of (Kirchler et al. 2021; Gustafson et al. 2023), and drawing from the ethical concerns raised by (Scheuerman, Pape, and Hanna 2021), we advocate for a move away from using visual cues to predict gender. The scholarly community, recognizing the inherent limitations and risks of such practices, increasingly endorses a non-prediction paradigm. This approach is not merely a technical consideration but a principled stand against reinforcing gender stereotypes and respecting the subjective experiences of gender identity.

The challenge is not only the technical limitations of gender classification but also the societal implications it carries. The field's exploration of fairness-through-blindness has revealed that attempts to resolve bias through classification can exacerbate the issue (Meister et al. 2023). The act of classification itself, particularly when it comes to gender, is fraught with the potential to solidify visual stereotypes rather than dismantle them. A clear illustration of the problem is the way non-conforming individuals or those who do not fit within the traditional gender binary may be labeled. These individuals can be classified as *Others*, and this othering can have harmful consequences. For example, if a system is designed to identify security threats based on deviations from gender norms, it could wrongly associate non-conformity with risk or danger (Mahalingam and Ricanek 2013; Scheuerman, Pape, and Hanna 2021). This doesn't just misclassify individuals; it actively contributes to a culture of discrimination and exclusion, wherein certain expressions of gender are viewed as suspicious or threatening. A more nuanced issue in this area is the intertwining of surveillance and hypervisibility of queer communities (Lingel 2021; Kalluri et al. 2023). The potential for these technologies to contribute to a culture of monitoring and control raises significant concerns, especially when considering the inclusivity of non-binary and transgender individuals. Is the inclusion of non-binary genders in CV technologies unequivocally beneficial, or does it inadvertently subject these identities to greater scrutiny and surveillance? This inquiry intersects with discussions on participatory approaches within CV (Birhane et al. 2022), suggesting pathways for queer communities to navigate their visibility and engagement with technology. Such approaches may offer the autonomy to opt-out or critically engage with CV applications, thereby mitigating the risks associated with unwarranted surveillance or misclassification.

Challenges, Open Questions and Future Directions

Computer Vision and Self-Perception: Metrics for Bias and Accuracy: The complexity of bias in CV is deeply interconnected with the constructs of self-perception and algorithmic interpretation. The current discourse confronts with the intricate relationship between algorithmic recognition and individual identity. The question of what defines accuracy—and consequently, bias—remains elusive, as there is no consensus on the standards for either within the field. This ambiguity in defining accuracy leads to profound difficulties in establishing metrics for measuring bias and accuracy. The challenge is not merely technical but philosophical, requiring a consensus that respects the myriad ways in which gender identity can be expressed and perceived. The endeavor to collect a dataset devoid of bias is challenging. To curate a dataset that genuinely reflects the vast array of gender identities requires a conscientious and methodical approach. It involves not only the gathering of data points but also a deep understanding and respect for the nuanced expressions of gender identity. Each data point must be examined for potential biases, and efforts must be made to prevent the perpetuation of stereotypes.

Unbiased Representation in Visual Datasets: Despite the efforts, the ingrained biases present in societal norms and the methodologies employed in data collection may prevent the attempts of creating such a dataset. These biases are pervasive and often ingrained in the structure of the data we seek to use for building inclusive and fair models. Moreover, the ambition to create an unbiased dataset must contend with the dynamism of social constructs. Gender identities evolve, and societal perceptions shift, rendering static datasets insufficient. A truly representative dataset, therefore, must be as fluid and adaptable as the identities it seeks to capture. It must evolve continually to mirror the changes in the way gender is understood and expressed. However, there is also an acceptance that must come with this pursuit. Complete neutrality may be an asymptote we approach but never fully reach. Recognizing this does not mean we abandon the effort; rather, we continuously strive to minimize bias, knowing that each step taken improves the integrity and inclusiveness of our data. This quest is iterative and reflective. It challenges us to question our assumptions, to listen to the communities represented, and to engage in a dialogue that shapes not only our datasets but also the very algorithms that will use them. The aim is not just to avoid reinforcing existing biases but to actively counteract them, creating datasets that foster diversity and promote equity.

Representation of Gender in Contexts of Uncertainty: Approaching gender representations in contexts of uncertainty demands a significant paradigm shift. The challenge lies in how systems can recognize and interpret the fluidity and spectrum of gender identities, which often do not conform to traditional binary norms. This uncertainty raises critical questions about the adaptability and sensitivity of algorithms. How can CV systems be designed to respectfully acknowledge and accurately represent the diversity of gender expressions, especially when these expressions don't align with clear-cut categories? Addressing this requires not only technological innovation but also a deeper understanding of gender as a social and cultural construct. It calls for collaborative efforts involving interdisciplinary expertise, including insights from gender studies, sociology, and ethics, to redefine the way gender is encoded and interpreted in computational models. This redefinition involves moving away from rigid classifications to more flexible, context-aware, and individual-centric approaches.

Beyond Fairness Rationalism: Traditionally, efforts to enhance fairness in AI systems have focused on identifying biases within training datasets and attempting to *debiased* these datasets, alongside refining algorithmic models with predefined mathematical definitions of fairness This approach presumes a rationalist perspective, assuming that optimally designed AI models will invariably be perceived as fair by all individuals. This presumption overlooks the inherently subjective nature of fairness, a social construct deeply influenced by individual perceptions, interpretations, and interactions with the world around us (De Cremer, 2020). In this sense, our review has shed light on a crucial aspect of AI fairness in the context of CV and our findings present an encouraging narrative: among the inclusive practices identified (Table 6), there is a balanced focus between technical adjustments (*Beyond Binary Classification, Blindness Approach, Gender as Discrete Variable*) and process-oriented practices (*Self-reported Gender, Informed Consent, Participatory Design*). Importantly, these process aspects focus on the subjective experiences of individuals, specifically relating to self-perception and self-definition of gender. These practices, while not directly addressing the subjectivity of fairness, are based on the recognition of the individual's role in defining their identity and how this self-definition intersects with the operation and outcomes of CV technologies. The insight that emerges from our review is profound: fairness beyond binarism and cisnormativity in CV extends beyond the technical domain to encompass the lived experiences and perceptions of those it seeks to represent. In summary, our research underscores the importance of integrating both technical and process-oriented practices in the development of CV systems. In detail, we advocate these two aspects:

- **Promote Participatory Design beyond Technical Solutions**: engage directly with non-binary, transgender and queer communities in the design and development phases of CV technologies. This approach ensures that the diverse needs and concerns of these communities are not only represented but actively inform the development process;
- **Conduct Longitudinal Studies on Social Impact**: undertake comprehensive longitudinal studies to evaluate the long-term social impacts of inclusive CV technologies on non-binary and queer individuals. Focus areas should include privacy, surveillance, and the overall representation of these communities.

Conclusion

This study represents a significant step in addressing gender bias in Computer Vision (CV), particularly from a non-binary perspective. It systematically reviews over 60 papers, revealing an increasing awareness and inclusion of non-binary genders in CV research, yet finding that practical applications remain fragmented and limited. The findings reveal that while there is growing acknowledgment of the complexities of gender, the translation of this understanding into practical CV applications is still in its infancy. The research critically evaluates existing approaches to gender in CV, highlighting the limitations of binary perspectives and showcasing emerging methodologies that challenge this viewpoint. This review underscores the importance of integrating gender theory into CV practices and calls for an interdisciplinary approach, particularly involving gender studies, to enrich the understanding and representation of gender in CV. It advocates for a comprehensive rethinking of gender representation in CV, encouraging the adoption of a more inclusive approach to gender identities.

List of Papers Included in the Review

(Akhyani et al. 2022) (Alasadi, Al Hilli, and Singh 2019) (Alasadi et al. 2020) (Alasadi et al. 2022) (Andrews et al. 2023b) (Andrews et al. 2023a) (Barlas et al. 2021) (Booth et al. 2021) (Cabello et al. 2023) (Cairns 2021) (Chin-Purcell and Chambers 2021) (Chong et al. 2021) (Díaz et al. 2023) (Dominguez-Catena, Paternain, and Galar 2023) (Friedrich et al. 2023) (Georgopoulos et al. 2021) (Gustafson et al. 2023) (Hamidi, Scheuerman, and Branham 2018) (Hazirbas et al. 2022) (Heldreth et al. 2023) (Herruzo 2021) (Hirota, Nakashima, and Garcia 2022) (Hoefsloot 2021) (Hoque et al. 2020) (Katzman et al. 2023) (Keswani and Celis 2021) (Kirchler et al. 2021) (Kumar et al. 2016) (Lee et al. 2023) (Leslie et al. 2022) (Lin, Kim, and Joo 2022) (Mahalingam and Ricanek 2013) (Mahalingam, Ricanek, and Albert 2014) (Mandal, Little, and Leavy 2023) (Meister et al. 2023) (Mousavi, Shahbazi, and Asudeh 2024) (Muthukumar, Kingsbury, and Mojsilović 2018) (Muthukumar et al. 2019) (Ovalle, Liang, and Boyd 2023) (Porgali et al. 2023) (Qi et al. 2022) (Quaresmini and Primiero 2023) (Ranjit et al. 2023) (Ravfogel et al. 2022) (Rosenbaum 2022) (Scheuerman and Brubaker 2018) (Scheuerman, Paul, and Brubaker 2019) (Scheuerman et al. 2020) (Scheuerman, Pape, and Hanna 2021) (Schumann et al. 2021) (Schumann et al. 2023) (Shen et al. 2023) (Smith et al. 2023) (Teotia, Lapedriza, and Ostadabbas 2022) (Thong, Joniak, and Xiang 2023) (Vijayan, Kareem, and Kizhakkethottam 2016) (Villalobos, Mery, and Bowyer 2022) (Wang et al. 2019) (Wang, Ramaswamy, and Russakovsky 2022) (WANG et al. 2023) (Weinschenk 2021) (Wu et al. 2020) (Xiang 2022)

References

Akhyani, S.; Abbasi, M.; Chen, M.; and Lim, A. 2022. Towards Inclusive HRI: Using Sim2Real to Address Underrepresentation in Emotion Expression Recognition. *arXiv preprint arXiv:2208.07472*. Supported by Huawei-SFU Visual Computing Joint Lab and NSERC, RGPIN/06908-2019.

Alasadi, J.; Al Hilli, A.; and Singh, V. K. 2019. Toward Fairness in Face Matching Algorithms. In *Proceedings of the 1st International Workshop on Fairness, Accountability, and Transparency in MultiMedia*, FAT/MM '19, 19–25. New York, NY, USA: Association for Computing Machinery. ISBN 9781450369152.

Alasadi, J.; AlHilli, A.; Atrey, P. K.; and Singh, V. K. 2022. A Generative Approach to Mitigate Bias in Face Matching using Learned Latent Structure. In *2022 IEEE Eighth International Conference on Multimedia Big Data (BigMM)*, 150–157.

Alasadi, J.; Arunachalam, R.; Atrey, P.; and ... 2020. A fairness-aware fusion framework for multimodal cyberbullying detection. *2020 IEEE Sixth* ....

Andrews, J.; Zhao, D.; Thong, W.; Modas, A.; and ... 2023a. Ethical considerations for collecting human-centric image datasets. *arXiv preprint arXiv ....*

Andrews, J. T. A.; Zhao, D.; Thong, W.; Modas, A.; and ... 2023b. Principlism Guided Responsible Data Curation. *dmlr.ai*.

Barlas, P.; Kyriakou, K.; Guest, O.; Kleanthous, S.; and Otterbacher, J. 2021. To "See" is to Stereotype: Image Tagging Algorithms, Gender Recognition, and the Accuracy-Fairness Trade-off. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW3).

Birhane, A.; Isaac, W.; Prabhakaran, V.; Diaz, M.; Elish, M. C.; Gabriel, I.; and Mohamed, S. 2022. Power to the People? Opportunities and Challenges for Participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '22. New York, NY, USA: Association for Computing Machinery. ISBN 9781450394772.

Booth, B. M.; Hickman, L.; Subburaj, S. K.; Tay, L.; and ... 2021. Bias and fairness in multimodal machine learning: A case study of automated video interviews. *Proceedings of the ....*

Butler, J. 1988. Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory. *Theatre Journal*, 40(4): 519–531.

Butler, J. 2006. *Gender Trouble: Feminism and the Subversion of Identity*. Routledge.

Cabello, L.; Bugliarello, E.; Br; l, S.; and Elliott, D. 2023. Evaluating Bias and Fairness in Gender-Neutral Pretrained Vision-and-Language Models. *arXiv preprint arXiv ....*

Cairns, S. 2021. *Defanging facial recognition: A statistical approach to bias mitigation and policy conditions for responsible use*. Ph.D. thesis, McGill University.

Chin-Purcell, L.; and Chambers, A. 2021. Investigating accuracy disparities for gender classification using convolutional neural networks. *2021 IEEE International ....*

Chong, T.; Maudet, N.; Harima, K.; and Igarashi, T. 2021. Exploring a Makeup Support System for Transgender Passing Based on Automatic Gender Recognition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. New York, NY, USA: Association for Computing Machinery. ISBN 9781450380966.

Collett, C.; and Dillon, S. 2019. AI and Gender: Four Proposals for Future Research. Technical report, institution = University of Cambridge.

Corbett-Davies, S.; Pierson, E.; Feller, A.; Goel, S.; and Huq, A. 2017. Algorithmic Decision Making and the Cost of Fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, 797–806. New York, NY, USA: Association for Computing Machinery. ISBN 9781450348874.

Dominguez-Catena, I.; Paternain, D.; and Galar, M. 2023. Gender Stereotyping Impact in Facial Expression Recognition. In Koprinska, I.; Mignone, P.; Guidotti, R.; Jaroszewicz, S.; Fröning, H.; Gullo, F.; Ferreira, P. M.; Roqueiro, D.; Ceddia, G.; Nowaczyk, S.; Gama, J.; Ribeiro, R.; Gavaldà, R.; Masciari, E.; Ras, Z.; Ritacco, E.; Naretto, F.; Theissler, A.; Biecek, P.; Verbeke, W.; Schiele, G.; Pernkopf, F.; Blott, M.; Bordino, I.; Danesi, I. L.; Ponti, G.; Severini, L.; Appice, A.; Andresini, G.; Medeiros, I.; Graça, G.; Cooper, L.; Ghazaleh, N.; Richiardi, J.; Saldana, D.; Sechidis, K.; Canakoglu, A.; Pido, S.; Pinoli, P.; Bifet, A.; and Pashami, S., eds., *Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 9–22. Cham: Springer Nature Switzerland. ISBN 978-3-031-23618-1.

Díaz, M.; Dev, S.; Reif, E.; Denton, E.; and Prabhakaran, V. 2023. SoUnD Framework: Analyzing (So)cial Representation in (Un)structured (D)ata. arXiv:2311.17259.

Fabbrizzi, S.; Papadopoulos, S.; Ntoutsi, E.; and Kompatsiaris, I. 2022. A survey on bias in visual datasets. *Computer Vision and Image Understanding*, 223: 103552.

Friedrich, F.; Schramowski, P.; Brack, M.; and ... 2023. Fair diffusion: Instructing text-to-image generation models on fairness. *arXiv preprint arXiv ....*

Galai, Y. 2023. Political Visual Literacy. *International Political Sociology*, 17(3): olad010.

Georgopoulos, M.; Oldfield, J.; Nicolaou, M. A.; and ... 2021. Mitigating demographic bias in facial datasets with style-based multi-attribute transfer. *... of Computer Vision*.

Gustafson, L.; Rolland, C.; Ravi, N.; Duval, Q.; Adcock, A.; Fu, C.-Y.; Hall, M.; and Ross, C. 2023. FACET: Fairness in Computer Vision Evaluation Benchmark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 20370–20382.

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, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356206.

Hazirbas, C.; Bang, Y.; Yu, T.; Assar, P.; Porgali, B.; and ... 2022. Casual Conversations v2: Designing a large consent-driven dataset to measure algorithmic bias and robustness. *arXiv preprint arXiv ....*

Heldreth, C. M.; Monk, E. P.; Clark, A. T.; Ricco, S.; and ... 2023. Which Skin Tone Measures are the Most Inclusive? An Investigation of Skin Tone Measures for Artificial Intelligence. *ACM Journal on ....*

Herruzo, A. 2021. Interactive Media in Public Spaces: Ethics of Surveillance, Privacy and Bias. *... Conference on Computer-Aided Architectural Design ....*

Hirota, Y.; Nakashima, Y.; and Garcia, N. 2022. Quantifying Societal Bias Amplification in Image Captioning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13450–13459.

Hoefsloot, M. 2021. *Automated Misgendering: an Inquiry Into the Ethics of Automatic Gender Recognition*. Ph.D. thesis, University of Twente.

Hoque, M. N.; Saquib, N.; Billah, S. M.; and Mueller, K. 2020. Toward Interactively Balancing the Screen Time of

Actors Based on Observable Phenotypic Traits in Live Telecast. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW2).

Kalluri, P. R.; Agnew, W.; Cheng, M.; Owens, K.; Soldaini, L.; and Birhane, A. 2023. The Surveillance AI Pipeline. arXiv:2309.15084.

Katzman, J.; Wang, A.; Scheuerman, M.; Blodgett, S. L.; Laird, K.; Wallach, H.; and Barocas, S. 2023. Taxonomizing and measuring representational harms: a look at image tagging. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press. ISBN 978-1-57735-880-0.

Keswani, V.; and Celis, L. E. 2021. Auditing for Diversity Using Representative Examples. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, 860–870. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383325.

Keyes, O. 2018. The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW).

Kirchler, M.; Graf, M.; Kloft, M.; and Lippert, C. 2021. Explainability Requires Interactivity. *arXiv preprint arXiv:2109.07869*.

Kumar, V.; Raghavendra, R.; Namboodiri, A.; and Busch, C. 2016. Robust transgender face recognition: Approach based on appearance and therapy factors. In *2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 1–7.

Lee, N.; Bang, Y.; Lovenia, H.; Cahyawijaya, S.; and ... 2023. Survey of Social Bias in Vision-Language Models. *arXiv preprint arXiv ....*

Leslie, D. 2020. Understanding bias in facial recognition technologies. *arXiv preprint arXiv:2010.07023*.

Leslie, D.; Katell, M.; Aitken, M.; Singh, J.; Briggs, M.; and ... 2022. Data Justice in Practice: A Guide for Developers. *arXiv preprint arXiv ....*

Lin, X.; Kim, S.; and Joo, J. 2022. FairGRAPE: Fairness-Aware GRAdient Pruning mEthod for Face Attribute Classification. In Avidan, S.; Brostow, G.; Cissé, M.; Farinella, G. M.; and Hassner, T., eds., *Computer Vision – ECCV 2022*, 414–432. Cham: Springer Nature Switzerland. ISBN 978-3-031-19778-9.

Lingel, J. 2021. Dazzle camouflage as queer counter conduct. *European Journal of Cultural Studies*, 24(5): 1107–1124.

Mahalingam, G.; and Ricanek, K. 2013. Is the eye region more reliable than the face? A preliminary study of face-based recognition on a transgender dataset. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 1–7.

Mahalingam, G.; Ricanek, K.; and Albert, A. M. 2014. Investigating the Periocular-Based Face Recognition Across Gender Transformation. *IEEE Transactions on Information Forensics and Security*, 9(12): 2180–2192.

Mandal, A.; Little, S.; and Leavy, S. 2023. Multimodal Bias: Assessing Gender Bias in Computer Vision Models with NLP Techniques. In *Proceedings of the 25th International Conference on Multimodal Interaction*, ICMI '23, 416–424. New York, NY, USA: Association for Computing Machinery. ISBN 9798400700552.

Meister, N.; Zhao, D.; Wang, A.; Ramaswamy, V. V.; Fong, R.; and Russakovsky, O. 2023. Gender Artifacts in Visual Datasets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 4837–4848.

Monea, A. 2022. 2. Straight Code. *The Digital Closet*.

Mousavi, M.; Shahbazi, N.; and Asudeh, A. 2024. Data Coverage for Detecting Representation Bias in Image Datasets: A Crowdsourcing Approach. In *27th International Conference on Extending Database Technology (EDBT)*. University of Illinois Chicago.

Muthukumar, V.; Kingsbury, B.; and Mojsilović, A. 2018. Understanding Unequal Gender Classification Accuracy from Face Images. *arXiv preprint arXiv:1812.00099*.

Muthukumar, V.; Pedapati, T.; Ratha, N.; Sattigeri, P.; Wu, C.-W.; Kingsbury, B.; Kumar, A.; Thomas, S.; Mojsilović, A.; and Varshney, K. R. 2019. Color-Theoretic Experiments to Understand Unequal Gender Classification Accuracy From Face Images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2286–2295.

Noble, S. U. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press. ISBN 9781479849949.

Ovalle, A.; Liang, D.; and Boyd, A. 2023. Should they? Mobile Biometrics and Technopolicy Meet Queer Community Considerations. *Proceedings of the 3rd ACM Conference on ....*

Page, M. J.; McKenzie, J. E.; Bossuyt, P. M.; Boutron, I.; Hoffmann, T. C.; Mulrow, C. D.; Shamseer, L.; Tetzlaff, J. M.; Akl, E. A.; Brennan, S. E.; Chou, R.; Glanville, J.; Grimshaw, J. M.; Hróbjartsson, A.; Lalu, M. M.; Li, T.; Loder, E. W.; Mayo-Wilson, E.; McDonald, S.; McGuinness, L. A.; Stewart, L. A.; Thomas, J.; Tricco, A. C.; Welch, V. A.; Whiting, P.; and Moher, D. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372.

Porgali, B.; Albiero, V.; Ryda, J.; and ... 2023. The Casual Conversations v2 Dataset. *... on Computer Vision ....*

Qi, Q.; Ardeshir, S.; Xu, Y.; and Yang, T. 2022. Fairness via Adversarial Attribute Neighbourhood Robust Learning. *arXiv preprint arXiv:2210.06630*.

Quaresmini, C.; and Primiero, G. 2023. Data quality dimensions for fair AI. *arXiv preprint arXiv:2305.06967*.

Ranjit, J.; Wang, T.; Ray, B.; and Ordonez, V. 2023. Variation of Gender Biases in Visual Recognition Models Before and After Finetuning. *arXiv preprint arXiv:2303.07615*.

Ravfogel, S.; Twiton, M.; Goldberg, Y.; and Cotterell, R. D. 2022. Linear Adversarial Concept Erasure. In Chaudhuri, K.; Jegelka, S.; Song, L.; Szepesvari, C.; Niu, G.; and

Sabato, S., eds., *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, 18400–18421. PMLR.

Rosenbaum, J. 2022. *AI perceptions of gender*. Ph.D. thesis, RMIT University.

Scheuerman, M.; and Brubaker, J. 2018. Gender is not a Boolean: towards designing algorithms to understand complex human identities. *In Participation+ Algorithms* ....

Scheuerman, M.; Pape, M.; and Hanna, A. 2021. Auto-essentialization: Gender in automated facial analysis as extended colonial project. *Big Data &Society*.

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. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW).

Scheuerman, M. K.; Wade, K.; Lustig, C.; and ... 2020. How we've taught algorithms to see identity: Constructing race and gender in image databases for facial analysis. *Proceedings of the ACM* ....

Schumann, C.; Olanubi, G. O.; Wright, A.; Jr, E. M.; and ... 2023. Consensus and Subjectivity of Skin Tone Annotation for ML Fairness. *arXiv preprint arXiv* ....

Schumann, C.; Ricco, S.; Prabhu, U.; Ferrari, V.; and Pantofaru, C. 2021. A Step Toward More Inclusive People Annotations for Fairness. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, 916–925. New York, NY, USA: Association for Computing Machinery. ISBN 9781450384735.

Schwemmer, C.; Knight, C.; Bello-Pardo, E. D.; Oklobdzija, S.; Schoonvelde, M.; and Lockhart, J. W. 2020. Diagnosing Gender Bias in Image Recognition Systems. *Socius*, 6: 2378023120967171.

Shen, X.; Du, C.; Pang, T.; Lin, M.; Wong, Y.; and ... 2023. Finetuning Text-to-Image Diffusion Models for Fairness. *arXiv preprint arXiv* ....

Smith, B.; Farinha, M.; Hall, S.; Kirk, H.; and ... 2023. Balancing the Picture: Debiasing Vision-Language Datasets with Synthetic Contrast Sets. *arXiv preprint arXiv* ....

Teotia, D.; Lapedriza, A.; and Ostadabbas, S. 2022. Interpreting Face Inference Models Using Hierarchical Network Dissection. ... *Journal of Computer Vision*.

Thong, W.; Joniak, P.; and Xiang, A. 2023. Beyond Skin Tone: A Multidimensional Measure of Apparent Skin Color. ... *on Computer Vision*.

Udefi, A. M.; Aina, S.; Lawal, A. R.; and Oluwarantie, A. I. 2023. An Analysis of Bias in Facial Image Processing: A Review of Datasets. *International Journal of Advanced Computer Science and Applications*, 14(5).

Vijayan, A.; Kareem, S.; and Kizhakkethottam, J. J. 2016. Face Recognition Across Gender Transformation Using SVM Classifier. *Procedia Technology*, 24: 1366–1373. International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015).

Villalobos, E.; Mery, D.; and Bowyer, K. 2022. Fair Face Verification by Using Non-Sensitive Soft-Biometric Attributes. *IEEE Access*, 10: 30168–30179.

WANG, A.; BAI, X.; BAROCAS, S.; and BLODGETT, S. 2023. Measuring stereotype harm from machine learning errors requires understanding who is being harmed by which errors in what ways.

Wang, A.; Liu, A.; Zhang, R.; Kleiman, A.; Kim, L.; and ... 2022. REVISE: A tool for measuring and mitigating bias in visual datasets. *International Journal of Computer Vision*, 130: 1790—-1810.

Wang, A.; Ramaswamy, V. V.; and Russakovsky, O. 2022. Towards Intersectionality in Machine Learning: Including More Identities, Handling Underrepresentation, and Performing Evaluation. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, 1–14. Seoul, Republic of Korea: ACM.

Wang, T.; Zhao, J.; Yatskar, M.; and ... 2019. Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations.

Weinschenk, K. 2021. Gendered Limits and Residuals: Computational Erasure of Indeterminate Modes of Being. *VIRGINIA JOURNAL OF GENDER STUDIES*, 1(2).

Wu, W.; Protopapas, P.; Yang, Z.; and ... 2020. Gender classification and bias mitigation in facial images. *Proceedings of the 12th* ....

Xiang, A. 2022. Being'Seen'vs.'Mis-Seen': Tensions between Privacy and Fairness in Computer Vision. *Harvard Journal of Law & Technology*, 36(1).