



**Data  
Governance  
Network**

Anchored by IDFC Institute

December 2021

Working Paper 20

# Cinderella's shoe won't fit Soundarya: An audit of facial processing tools on Indian faces

*Gaurav Jain and Smriti Parsheera*





---

## Data Governance Network

The Data Governance Network is developing a multi-disciplinary community of researchers tackling India's next policy frontiers: data-enabled policymaking and the digital economy. At DGN, we work to cultivate and communicate research stemming from diverse viewpoints on market regulation, information privacy and digital rights. Our hope is to generate balanced and networked perspectives on data governance – thereby helping governments make smart policy choices which advance the empowerment and protection of individuals in today's data-rich environment.

### About Us

Gaurav Jain is a policy professional with training as an engineer and is currently a sector economist at the International Finance Corporation. He has experience working across governments, research organisations and development finance institutions, including with the University of Chicago Trust in India and the European Bank for Reconstruction and Development. He has a Master in Public Policy from the University of Oxford and a B.Tech. (Hons.) from the Indian Institute of Technology, Kharagpur.

Smriti Parsheera is a lawyer and public policy researcher. She is a Fellow at the CyberBRICS Project, FGV Law School and a PhD candidate at IIT, Delhi. She has previously worked with the National Institute of Public Finance and Policy (NIPFP), the Competition Commission of India, the UNDP and the Financial Sector Legislative Reforms Commission. She studied law at National Law School, Bangalore followed by a LLM from the University of Pennsylvania.

Jain and Parsheera were previously researchers with the technology policy group at NIPFP, a centre for research in public economics and policies.

### Disclaimer and Terms of Use

The views and opinions expressed in this paper are those of the authors and do not necessarily represent those of the organisation.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

### Design

Cactus Communications

### Suggested Citation:

Gaurav, J., Parsheera, S. (2021). Cinderella's shoe won't fit Soundarya: An audit of facial processing tools on Indian faces. Data Governance Network Working Paper 20.

---

## **Abstract**

The increasing adoption of facial processing systems in India is fraught with concerns of privacy, transparency, accountability, and missing procedural safeguards. At the same time, we also know very little about how these technologies perform on the diverse features, characteristics, and skin tones of India's 1.34 billion plus population. In this paper, we test the face detection and facial analysis functions of four commercial facial processing tools on a dataset of Indian faces. The tools display varying error rates in the face detection and gender and age classification functions. The gender classification error rate for Indian females faces is consistently higher compared to that of males – the highest female error rate being 14.68%. In some cases, this error rate is much higher than that shown by previous studies for females of other nationalities. Age classification errors are also high. Despite taking into account an acceptable error margin of plus or minus 10 years from a person's actual age, age prediction failures are in the range of 14.3% to 42.2%. These findings point to the limited accuracy of facial processing tools, particularly for certain demographic groups, and the need for more critical thinking before adopting such systems.

---

‡We thank Renuka Sane, Ajay Shah, Noopur Raval, Vidushi Marda, Robin Zachariah Tharakan, Rishab Bailey, Vimal Balasubramaniam and Devendra Damle for valuable discussions and comments. Earlier drafts of this paper were discussed in meetings of the Data Governance Network, in a session on 'AI Bias Beyond the Western Lens: Perspectives From India' organised by the authors at Mozilla Festival, 2021 and at the CVPR 2021 Workshop on 'Beyond Fairness: Towards a Just, Equitable, and Accountable Computer Vision' organised by Timnit Gebru and Emily Denton. We are grateful to the participants at these events for inputs that helped us in improving the drafts. All errors are our own.

---

# **Table of Contents**

<b>1. Introduction</b>	<b>04</b>
<b>2. Related Work</b>	<b>05</b>
<b>3. Methodology</b>	<b>07</b>
3.1 ECI Faces Dataset	07
3.2 Facial Processing Tools	08
3.3 Method of Analysis	08
<b>4. Ethical Considerations</b>	<b>09</b>
<b>5. Findings and Implications</b>	<b>11</b>
5.1 Face Detection	11
5.2 Gender Clasification	12
5.3 Age Prediction	14
<b>6. Conclusion</b>	<b>15</b>
<b>References</b>	<b>17</b>

---

# **1. Introduction**

India is seeing a rise in the use of artificial intelligence (AI) based systems, with facial processing being one of the most common use cases. The AI Observatory contains a database of 65 automated decision-making systems in India, one-third of which relate to the use of facial recognition (Joshi, 2020). Project Panoptic, another useful tracker developed by the Internet Freedom Foundation, identifies 75 facial recognition projects (Jain et al., 2020). The applications of this technology, in the public and private sectors, include purposes such as law enforcement, employment screening, device security, attendance systems, and know your customer checks (Parsheera, 2019). Amidst the COVID-19 crisis, government authorities also turned to testing facial authentication as one of the possible means of identity verification for accessing vaccines (Barik, 2021).

India has a fairly rich body of literature documenting the concerns raised by the growing use of facial processing – on grounds of lack of transparency, privacy concerns, biased outcomes, and a range of structural problems (Vipra, 2021; Bhandari, 2021; Jain, 2020; Aneja and Chamuah, 2020; Kovacs, 2020; Parsheera, 2019; Marda, 2019a; Basu and Sonkar, 2019). But empirical work on this subject still remains limited. This includes a dearth of studies auditing the performance of facial processing tools on the diversity of Indian faces, a gap that we aim to bridge through this paper.

While focusing on accuracy of facial processing, specifically in terms of measurement of error rates, we would like to emphasise that the accuracy is a necessary but not sufficient condition for the adoption of this technology (Parsheera, 2019). Notably, even with 100 percent accuracy facial processing systems would still be fraught with concerns of privacy and accountability. In fact, as noted by Kalluri (2020), a more accurate facial recognition system only becomes a more potent weapon in the hands of exploitative companies and oppressive state agencies. Yet, the developers and adopters of facial processing technologies (FPT) continue to rely on claims of increased performance and accuracy to push for greater adoption of these systems. A critical evaluation of how well FPT actually work, and for whom, therefore, becomes important.

In this paper, we audit the face detection and analysis functions of four commercially available FPT tools – Microsoft Azure’s Face, Amazon’s Rekognition, Face++, and FaceX – on a dataset of Indian faces. We do this using publicly available images of election candidates sourced from the website of the Election Commission of India (ECI). The goal is to understand how these tools perform in carrying out face detection, gender classification and age estimation functions on Indian faces, and the likely implications of errors in the results.

Accuracy of facial processing is a highly context-specific metric. Its outcomes tend to vary based on the quality of images being used, the nature of the use case, and the characteristics of the demographic population. The diversity of Indian faces, consisting of over 1.34 billion individuals from a mix of racial, cultural, genetic, and environmental backgrounds, therefore, serves as a fertile ground for questioning the accuracy of facial processing systems. Given this diversity of the Indian population, it would be futile, and even erroneous, to try and classify ‘Indians’ as a distinct racial category (Khan and Fu, 2021). Accordingly, the references to ‘Indian faces’ in this paper do not imply a racial categorisation but simply refer to the diversity of facial features exhibited by people living in the territory of India.

While espousing the need for research of this nature, we are also mindful of the sensitive nature of facial data and the ethical considerations that ought to guide this field of research. We, accordingly, focus on understanding the literature on ethical use of publicly available datasets (Buchanan and Zimmer, 2021 ;

---

Vitak et al., 2016) and have suitably tried to apply the best practices on assessment of benefits and harms in the design and implementation of our work.

The rest of the paper is organised as follows. Section two presents a review of related work on auditing of FPT tools and Indian face datasets. Section three outlines the methodology, including a description of our dataset and the selected methods of analysis. Section four discusses the ethical considerations that motivated the design of this study. Section five presents the findings of our research and their implications. We conclude in section six with a brief summary of the findings.

## **2. Related work**

We situate this work in the field of algorithmic auditing studies, particularly those that have focused on the performance of commercially available off-the-shelf facial processing tools (Khalil et al., 2020; Raji, Gebru et al., 2020; Buolamwini and Gebru, 2018). The Gender Shades study by Buolamwini and Gebru has been pioneering work in this field. Using a database of images of parliamentarians from African and European countries, the authors reviewed the performance of the gender classification function of the FPT tools offered by Microsoft, IBM and Face++.<sup>1</sup> They found that the tools generated a disproportionately higher gender classification error, as high as 34.7%, for darker-skinned females as compared to just 0.8% for lighter-skinned males (Buolamwini and Gebru, 2018). Subsequent studies have extended a similar methodology to other tools like Amazon, Kairos, and Clarifai (Raji and Buolamwini, 2019; Scheuerman et al., 2019).

Jung et al. (2018) also use a similar approach to infer the accuracy of gender, race, and age attribution features using multiple publicly available image datasets. But they do not relate the errors to variation in gender or skin tones. The authors of this study found that the tools offered by Microsoft, IBM, Face++ and Amazon were ‘generally proficient at determining gender’ but all the tools performed poorly on inferring age (Jung et al., 2018). A collective reading of these studies points to the importance of a granular exploration of the intersectionality of factors such as race, gender and age while studying AI bias.

The studies referred to above deal with different aspects of *facial processing*, which is broader in scope than facial recognition. Facial recognition refers to a one-to-one verification or one-to-many identification by matching two or more images (Romine, 2020). But facial processing also includes the functions of face detection (identifying a face in an image) and facial analysis (inferring characteristics such as gender, ethnicity, and emotions) (Raji, Gebru et al., 2020). Although our paper focuses on facial processing functions other than recognition, studies have found that the accuracy of facial recognition also tends to vary across demographic groups (Klare et al., 2012; Snow, 2018; Grother et al., 2019). Instances of poor performance of facial recognition have also come to light in real world applications of the technology in India. As per admissions made by government agencies before the Delhi High Court, use of facial recognition for identification of missing persons had accuracy rates as low as 2 percent (PTI, 2018). Further, the software often resulted in the matching of pictures of missing boys as girls (PTI, 2019).

Some of the literature referred to above has been instrumental in highlighting the limitations and pitfalls of facial processing. But just as much of the development of FPT is taking place in contexts far

---

<sup>1</sup>The Pilot Parliaments Benchmark constructed by the authors for this purpose consisted of 1,270 images of parliamentarians from three African countries (Rwanda, Senegal, South Africa) and three European countries (Iceland, Finland, Sweden).

removed from the Indian one, the critical discourse on facial processing is also coloured by western datasets and institutional perspectives (Marda and Narayan, 2020; Sambasivan et al., 2021). It has been argued that the conventional or western-centric model of algorithmic fairness can not be directly applied to an Indian context due to different axes of discrimination in India, which encompasses factors like caste, religion, ethnicity and class (Sambasivan et al., 2021). Further, differences in legal context, societal perspectives, state capacity, and institutional structures are also critical to understanding the politics and governance of AI systems (Marda, 2019b; Parsheera, 2021). In line with the calls for more context specific research on facial processing, this paper examines how available commercial FPT tools perform on Indian faces and the implications of inaccurate results for different demographic groups.

To the best of our knowledge, there has not been any empirical work on issues of accuracy or fairness focusing on Indian faces. Existing studies on facial processing in India can broadly be divided into two streams. The first consists of technical studies that focus on the applications and capabilities of facial processing. This covers issues such as classification of faces based on regional affiliations – North and South Indian (Katti and Arun, 2019) or North, East and South Indian (Sarin and Panda, 2020), identification of genetic disorders in children (Narayanan et al., 2019), and detection of emotions (Singh and Benedict, 2020). The second stream of work consists of research papers, reports and other critical perspectives on the use of facial processing, highlighting associated risks, harms, and modes of regulation (Bhandari, 2021; Joshi, 2020; Jain, 2020; Aneja and Chamuah, 2020; Kovacs, 2020; Parsheera, 2019; Marda, 2019a; Basu and Sonkar, 2019). We locate this paper in this latter stream of work but distinguish it from the existing studies in terms of the use of quantitative methods of analysis.

One of the reasons why empirical research auditing the performance of facial processing in India remains limited could be due to the scarcity of appropriate image datasets. Table 1 outlines the key features of the few India-specific facial datasets that are publicly available. As the table shows, most of these datasets cover only a limited number of unique individuals. Further, covered subjects are often university students, volunteers from major cities (Lakshmi et al., 2021; Gupta et al., 2019; Happy et al., 2017) or movie celebrities (Setty et al., 2013), which limits their scope in terms of age, regional diversity and urban-rural representation. In addition to these, some Indian faces also find a place in popular fair computer vision datasets but are generally grouped under the broader category of South Asian faces (Khan and Fu, 2021). Realising the limitations of existing Indian faces datasets for meeting our objectives, we created a new dataset of facial images – the ECI Faces Dataset – using the methodology described below.

Table 1: Existing databases of Indian faces

Dataset	Faces	Individual	Age Range	Remark
NEI-DB (Saha et al., 2013)	59,850	630	Not Recorded	Covering 5 North Eastern states (both tribal and non-tribal faces)
CFD-India (Lakshmi et al., 2021)	142	142	18-50 Years	Volunteers from Delhi, representing different regions of India
IIITM-G (Gupta et al., 2019)	1,928	107	Not Recorded	Students and staff of the ABV IIITM-G
IMFDB (Setty et al., 2013)	34,512	100	Broad Estimates	Screenshots from 103 movies (5 languages); age is estimated
IFAD (Sharma and Patterh, 2015)	3,296	55	Not Available	Screenshots of Indian Hindi movies, age not estimated
ISED (Happy et al., 2017)	428	50	18-22 Years	Students from IIT Kharagpur, representing different regions of India



## 3. Methodology

We created a new dataset using publicly available facial images from the ECI’s candidate affidavit portal.<sup>2</sup> This section presents a description of the ECI Faces Dataset and the manner in which it was processed. We also describe the criteria used for the selection of the chosen FPT tools and the methods of analysis that were deployed.

### 3.1 ECI Faces Dataset

The ECI’s candidate portal contains photographs and other information submitted by electoral candidates in their affidavits. We gathered 49,346 observations of candidates who filed nominations to contest in the Parliamentary and State Assembly elections (both general and bye elections) held between May 2019 to Nov 2020. The data cleaning process was done in the following steps.

1. Duplicate entries with the same image were removed.
2. All entries with missing data or data entry errors, such as the candidate’s age being described as zero, were removed.
3. A manual check was done for multiple images of the same person after filtering the results for matching candidate and father/husband’s name.
4. Images which couldn’t be processed by any one of the tools were removed. This resulted in the removal of 15 images. Notably, none of these images presented an error across all the tools, indicating tool specific image processing limitations.

Following this process, we arrived at a dataset of 32,184 unique observations, consisting of a photograph and details like State, age and gender for each entry. We, retained all valid data entries even though the ECI may have subsequently rejected some of these nominations for other reasons. For instance, this would include nominations filed by candidates below 25 years of age, which is the minimum age to stand for elections in India.<sup>3</sup>

These 32,184 observation cover all the constituencies of India, which means that persons from all States and Union Territories (UT) are represented in the dataset. The ECI allows persons to select their gender as male, female, or third gender. In line with a 2014 decision by the Indian Supreme Court, the term third gender includes persons belonging to the hijra community and transgender persons (National Legal Services Authority vs Union of India, 2014) The dataset consists of 3,524 female (10.95%), 28,646 male (89.01%) and 14 third gender (0.04%) candidates. Their ages range from 18 – 100 years (see Table 2).

Table 2: Descriptive statistics about the dataset

Variable	N	Unique N	Percentage
State/UT	32,184	36	-
<b>Gender</b>			
Female	3524	-	10.95%
Male	28,646	-	89.01%
Third Gender	14	-	0.04%
<b>Age Brackets</b>			
18-25 Years	55	-	0.17%
25-40 Years	10709	-	33.27%
40-55 Years	13936	-	43.3%
55-70 Years	6636	-	20.62%
70-85 Years	833	-	2.59%
85-100 Years	15	-	0.05%

<sup>2</sup> Available at <https://affidavit.eci.gov.in/candidate-affidavit>, accessed on 1 December 2020.

<sup>3</sup> Articles 84(b) and 173 (b), Constitution of India.

---

The data composition is influenced by various structural factors, like gender, caste, class, and religion, that shape who files an election nomination in India. Further, Some States also held assembly elections in the period of data collection, leading to additional representation in the dataset compared to States and UT covered only in the 2019 Parliamentary election. But, despite these limitations, the dataset is unique in its coverage and size and, therefore, forms a suitable basis for this research.

Most of the images in the dataset adhere to the Indian government's face image quality requirements for e-governance applications (GoI, 2010). The standards require a front-facing image showing the face and shoulders of the person, with uniform lighting against a plain white or off-white/blue background. The photographs on the ECI's portal could be digital images uploaded directly by the candidate or scanned versions of the physical photograph submitted with the affidavit. The submitted information is also verified by a designated returning officer (ECI, 2020a; ECI, 2020b), ensuring a certain level of manual validation check before the images are made available on the portal.

### **3.2 Facial processing tools**

The selection of the commercial FPT tool was done taking into account the tools covered in previous audit studies and the availability of gender and age classification features in those tools. Accordingly, we identified three tools, Microsoft Azure's Face, Amazon's Rekognition, and Face++, as being the most suitable for our purposes. We also added one tool of Indian origin - FaceX. This was done keeping in mind the studies that indicate that algorithms from a particular region tend to have comparatively more accurate results for images of persons belonging from that region (Grother et al., 2019; O'Toole, 2011).

### **3.3 Method of analysis**

In line with previous auditing studies (Jung et al., 2018; Buolamwini and Gebru, 2018; Raji and Buolamwini, 2019), we focus on studying the errors in the Application Programming Interface (API) responses generated by the tools. We assess the accuracy of the predicted results by comparing the generated results with the candidate's self-disclosed gender and age information, which is treated as the ground truth. Two of the selected tools – Amazon and FaceX – generated an age range and not a value. In these cases, the mean of the provided range limits was used as the predicted age. But since FaceX provides age predictions from a set of pre-determined age ranges, we recognise that this may lead to higher error rates compared to other tools that either provide an age value or a dynamic age range.

The tools also vary in the presentation of their responses on gender classification, Microsoft and Face++ only mention the predicted gender without providing a probability for the prediction. On the other hand, Amazon and FaceX provide a probability score to indicate how sure the model is of the predicted gender. To enable comparison across tools, we have not taken into account the probability scores in our analysis. Further, in line with the terminology used by the ECI's affidavit portal and the explanation provided by the selected tools, we use the term gender, as opposed to sex, while discussing the findings and implications. This is also consistent with the individual's ability to self identify their gender while filing the election papers.

As noted in the previous section, there are certain other grounds of marginalisation, like a person's caste and tribal status, that are critical to studying fairness in the Indian context. However, this information is not available in the candidate information provided on the ECI's affidavit portal. While there are example of studies that have inferred caste details based on the last names of individuals (Fisman et al., 2017; Bhagavatula et al., 2019), we find that these linkages are prone to errors. For instance, a given last name could correspond to multiple castes or tribes with variations based on the person's religion or list

---

persons belonging to those groups. Our findings, therefore, do not include observations on caste or tribe based categorisation of Indian faces although the implications of incorrect attribute classification are likely to be all the more stark for persons belonging to these marginalised groups.

## **4. Ethical considerations**

Given the sensitivity of facial biometric data and its intrusive applications, research on automated facial analysis raises several ethical concerns. Examples of problematic research studies include the use of ethnicity recognition tools to identify Chinese Uyghur Muslims (C. Wang et al., 2019) and predicting sexual orientation using images from dating and social media sites (Y. Wang and Kosinski, 2018). Such studies have been criticised for enabling the surveillance and prosecution of vulnerable groups (Noorden, 2020; Moreau, 2019) and violating expectations of privacy and sensitivity of the content (Sweeney, 2017). Besides being ethically problematic, researchers have highlighted that several use cases of facial processing, such as for emotion recognition, also rest on unsound technical and scientific foundations (Marda and Ahmed, 2021).

While they may not generate the same types of concerns as applied AI research, algorithmic auditing studies are also bound by various ethical considerations. These considerations revolve around the source of the image dataset being used for the audit, its representativeness, and the rights of the covered individuals. Another concern that has been highlighted is that such studies may overemphasise the value of demographic representation in datasets rather than questioning the political or institutional context in which the technology is being used (Hoffmann, 2019).

Accordingly, we find it important to highlight some of the key ethical concerns that have informed this research. But, before doing so, we must emphasise that the scope of this paper is focused on testing the detection and classification of publicly available images by commercial FPT tools. We have not used the available images for developing or training any algorithms or applied the tools for any particular use case. The methodology and objective of our work is therefore very different from the problematic types of applied AI studies referred to above.

The first issue that we had to address while designing this study related to that of individual consent. Obtaining consent is generally not regarded as necessary for research relying on publicly available data (Buchanan and Zimmer, 2021). India's Personal Data Protection Bill, 2019, which is still in a draft stage, also recognises the 'processing of publicly available personal data' to be a reasonable purpose that can be pursued without consent from the individual.<sup>4</sup> This is in addition to the proposed exemption for research purposes.<sup>5</sup> Such an approach has, however, been questioned by some, particularly in the context of 'public' data gleaned from social media sites (Zimmer, 2010; Sweeney, 2017; Ravn et al., 2020). Relying on Helen Nissenbaum's conception of 'contextual integrity', Vitak et al. (2016) note that while obtaining informed consent may not always be possible for online datasets, researchers should respect 'the norms of the contexts in which online data was generated'.

The data on nomination affidavits filed by electoral candidates, which forms the basis of this research, is put out by the ECI for public information and transparency purposes. The ECI's affidavit portal does not contain any restrictions on the download or use of its data. On the contrary, the ECI has clarified, the

---

<sup>4</sup> Section 14(2)(g), Personal Data Protection Bill, 2019.

<sup>5</sup> Section 38, Personal Data Protection Bill, 2019.

---

“Affidavits and counter-affidavits are available for citizens to view and download” (ECI, 2019). Further, it can be presumed that the concerned persons have taken a conscious decision to be part of public life and put out their information for that purpose. While that does not imply an approval for being part of a facial processing audit, mandatory data put out by public figures arguably stands on a different footing from user-generated data on social media sites, which has more often been the subject of data ethics concerns. Taking into account the context of the data and the safeguards described below, we chose to proceed with the research without seeking the consent of individuals whose images were covered in the dataset.

We adopted the following safeguards to minimise any direct harm to the individuals and prevent any other potential negative societal implications.

First, following the principle of data minimisation, we limited the data collection exercise only to the fields that were necessary to give effect to our research methodology. Details like political party affiliations and financial disclosures that are contained in the candidate affidavits but were not relevant to our study were not captured in the dataset.

Second, we selected a period of study during which a General Election had been held in the country. This ensured the inclusion of election candidates from all parts of the country in the dataset hence enhancing the representativeness of the data. While this data is still not fully representative of the population of India, it certainly performs better on this front compared to other publicly available Indian facial datasets.

Third, we excluded the evaluation of caste or tribe affiliations as inference of these affiliations from other available pieces of information like names or constituencies was fraught with the risk of falsely indicating the presence or absence of bias against particular groups on account of methodological limitations.

Fourth, in order to protect the privacy of the individuals, we have chosen not to disclose any personally identifiable information in the paper. This includes abstaining from the use of any illustrative images while presenting the results.

Fifth, we decided not to put out the gathered image dataset in the public domain. While this information is already available to the public, our decision to not release the dataset is motivated by the desire to avoid aiding potentially undesirable secondary uses of the cleaned and collated dataset. We have, however, released the code used to get the API responses from the selected tools and the results generated by the processing of the images.<sup>6</sup> This can be used for establishing the verifiability of the results.

In light of these safeguards, we believe that the benefits of this research, in terms of furthering the agenda of algorithmic accountability by generating evidence on the performance of FPT tools on Indian faces, outweigh any potential harms. While our objective is to further the agenda of algorithmic accountability, it is possible that the results may also be used by entities trying to make a case in favour of facial processing on the ground that the technology seems to work ‘in most cases’. We aim to mitigate this possibility by highlighting the serious implications of failed face detection and misclassification in the next section of the paper.

---

<sup>6</sup>The code can be accessed at <https://github.com/gauravrpjain/auditing-fpt-eci-faces-db>

## 5. Findings and implications

This section presents the results of our audit of the performance of the selected tools in carrying out the functions of face detection, gender classification and age estimation. We also discuss the implications of the observed errors taking into account some of the scenarios in which such functions may be deployed. This includes the problematic use case of deployment by police and investigative agencies for law enforcement purposes. While companies like Microsoft and Amazon have currently adopted bans and moratoriums on the supply of facial recognition technology to police authorities, there are numerous others that continue to do so (Greig, 2021). The use case specific discussions below, therefore, remain relevant even though the policies and detection and classification error rates may vary for different tools.

### 5.1 Face detection

Since each image in the ECI Face Database is a headshot, face detection is considered to be a success if the tool is able to detect one face in the image. Detection error

(D) represents a failure in detecting a face.

$$D = \begin{cases} 0; & \text{if } n_{faces} = 1 \\ 1; & \text{if } n_{faces} \neq 1 \end{cases}$$

where  $n_{faces}$  is the number of faces detected. Table 3 reports the mean detection error for each tool and Table 4 provides the gender break-up of the successful cases of face detection. Only the successful cases of face detection are used for further analysis of the gender classification and age prediction functions.

Table 3: Error Rates in face detection across FPT tools

Tool	N	Faces Detected			Overall
		Zero	Multiple	One	
Amazon	32,184	7	9	32,168	0.05%
Face++	32,184	586	2	31,596	1.83%
FaceX	32,184	649	213	31,319	2.69%
Microsoft	32,184	1,018	3	31,163	3.17%

Table 4: Successfully detected faces for facial analysis

Tool	No.	Female (F)	Male (M)	Third (T)
Amazon	32,168	3,522	28,632	14
Face++	32,596	3,480	28,102	14
FaceX	32,319	3,475	27,830	14
Microsoft	32,163	3,438	27,711	14

The four tools differ significantly in their ability to detect faces. Amazon performed the best, with an error rate of 0.05% while Microsoft performed the worst, reporting an error rate of 3.17%. To contextualise this in terms of numbers, out of the 32,184 persons in the dataset, Microsoft was unable to accurately detect the faces of a little over a thousand persons. This figure was similarly high for FaceX, which failed to detect 865 of the faces. While significant in themselves, the implications of these figures become all the more relevant when we take into account a population size of over 1.34 billion people, with many individuals being forced to encounter multiple facial processing projects.

The detection of a face in an image or live video is the logical starting point in any facial analysis process. Face detection could sometimes serve as an end in itself, for instance in the auto-focus feature of a camera, object detection by autonomous vehicles, or while estimating the size of a crowd. In most other cases, detection is followed by other types of processing, such as classification, emotion detection, identification or verification. Failure to have one’s face detected, therefore, implies the system’s inability to carry out any further processing, hence leading to the potential exclusion of the individual.

This may be regarded as a favourable outcome in some cases, like exclusion from detection by surveillance cameras. However, the ubiquitous adoption of facial recognition in multiple contexts means that non-detection could also impose significant inconvenience and costs. This can be particularly troubling when face recognition is used for purposes like access control, conducting know your customer checks, entry into airports and other buildings, delivery of welfare schemes and vaccine delivery (Manikandan and Shukla, 2020; Chakrabarty, 2020; Chandna, 2021; Lalwani, 2019). Many of these services may offer an alternate route of manual human verification but availing the same could be subject to significant inconvenience, delays, and costs. Experience from the Aadhaar project also shows that manual alternatives to biometric authentication failures often do not work in practice, which could lead to the exclusion of legitimate beneficiaries (Khera, 2019).

## 5.2 Gender classification

Gender classification is considered to be a success if the predicted gender is the same as the self-described gender. The binary gender treatment of all tools, which assumes gender to be physiologically based and immutable (Keyes, 2018; Scheuerman et al., 2019), makes the tools completely incapable of accurately classifying the gender of the fourteen persons belonging to the third gender in the dataset.<sup>7</sup> For the remaining cases (where the self-described gender is either male or female) gender classification error (G) represents a failure in classifying gender.

$$G = \begin{cases} 0; & \text{if } \hat{g} = g; n_{\text{faces}} = 1 \\ 1; & \text{if } \hat{g} \neq g; n_{\text{faces}} = 1 \end{cases}$$

where  $\hat{g}$  is predicted gender;  $\hat{g} \in (F, M)$  and  $g$  is self described gender;  $g \in (F, M)$ .

Table 5: Error rates for gender classification across FPT tools

Tool	No.	Overall	Female (F)	Male (M)	F-M	t-test
Amazon	32,154	0.66%	2.56%	0.42%	2.13%	7.94***
Face++	31,582	1.85%	14.68%	0.26%	14.42%	24.01***
FaceX	31,305	2.36%	10.47%	1.35%	9.12%	17.41***
Microsoft	31,149	0.21%	0.79%	0.13%	0.65%	4.28***

Note: P<0.1; \*P<0.05; \*\*P<0.01; \*\*\*P<0.001

<sup>7</sup> As per the documentation provided by Microsoft about its Azure tool, the possible values for gender are male, female, and genderless. However, there has been no cases of a face being classified as genderless in literature as well as in our dataset were classified as genderless.

Table 5 reports the mean gender classification error rate for each tool. It shows that all the tools consistently fare poorer in the classification of female faces. This is particularly true in the case of Face++, which has a near-zero error rate for males but classifies 14.68% of females as males. Face X and Amazon also reflect a noticeable gender error gap (difference between error for females and males) of 9.12% and 2.14%, respectively. These gender error gaps are statistically significant for all the tools. Figure 1 shows how the gender classification error varies across age for different gender groups. It reflects that the chances of misclassification are consistently higher for females compared to males across all age groups.

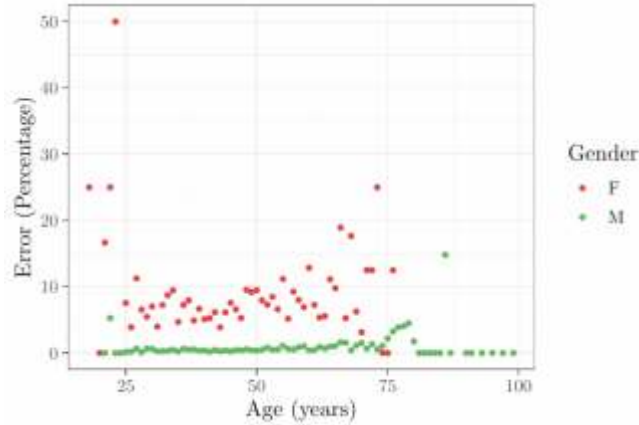


Figure 1: Gender classification error rates across age groups

Of the selected tools, Amazon, Face++ and Microsoft have previously been covered by other gender classification studies (Buolamwini and Gebru, 2018; Raji and Buolamwini, 2019). A comparison of our results with previous studies presents some interesting findings, particularly in the case of the Chinese company, Face++ (Table 6). Our study finds error rates for Face++ to be as high as 14.68% compared to an error rate of 2.5% for females in the study by Raji and Buolamwini (2019). The error figures are more comparable for Amazon and Microsoft. This could possibly be an indication of the differences in the representation of Indian faces in the training dataset used by different firms and the relative importance of the Indian market for them. Further, the improved performance of Face++ between the 2018 and 2019 studies, which are based on the same data set (the Pilot Parliaments Benchmark) points to the tendency to introduce band-aid measures in response to specific studies while leaving the door open for other problems and inaccuracies.

Table 6: Gender classification error rates for Face++ across studies

Study	F	M	F-M
Buolamwini and Gebru (2018)	21.3%	0.7%	20.6%
Raji and Buolamwini (2019)	2.5%	0.9%	1.6%
Our study (2021)	14.68%	0.26%	14.42%

Some examples of automated gender classification include access control for genderdemarcated spaces and targeted advertisements (Keyes, 2018; Leufer, 2021). Gender could also become the basis for automatic filtering of results for law enforcement purposes, such as identification of a suspect or a missing person based on their gender. Misclassification in such cases could result in incorrect identification, search or tracking of individuals. Similarly, use of gender analysis tools while filtering job applications could lead to unfair exclusions. The presence of humans in decision making chains could perhaps minimise some of these harms, but there is limited transparency and accountability about the oversight mechanisms incorporated by different users of such systems.

Finally, the use of facial processing could also have an impact on other models that do not explicitly include gender as a parameter for decision making. Research has shown a gendered bias to be present in all sorts of algorithmic systems from recruitment to text embeddings and even hate speech detectors (Xia et al., 2020; De-Arteaga et al., 2019; Prates et al., 2019; Sap et al., 2019; Dixon et al., 2018; Parsheera, 2018). Many of these studies illustrate how algorithms can acquire a gender bias even when gender is not an explicit input. Any model built using FPTs for purposes beyond facial analysis could also implicitly extract gender information and contribute toward amplifying existing biases.

### 5.3 Age prediction

Recognising the challenges of estimating a person’s exact age, we relied on age bins or brackets to calculate the age classification errors. Age prediction is considered a success if the difference between the actual age and the person’s self-described age falls within the threshold limits, of 2, 5 and 10 years. This creates acceptable age bins of 5, 11 and 21 years, respectively. Age prediction error ( $A$ ) represents a failure to predict age successfully.

$$A = \begin{cases} 0; & \text{if } |\hat{a} - a| \leq t; n_{\text{faces}} = 1 \\ 1; & \text{if } |\hat{a} - a| > t; n_{\text{faces}} = 1 \end{cases}$$

where  $\hat{a}$  is predicted age,  $a$  is self described age, and  $t$  is threshold limit;  $t \in \{2, 5, 10\}$ .

Table 7: Error rates in age prediction across FPT tools

Tool	No.	Error Rate - Age Bins		
		5 Year	11 Year	21 Year
Amazon	32,168	80.56%	60.36%	34.72%
Face++	32,596	77.85%	54.5%	25.34%
FaceX	32,319	83.48%	65.8%	42.19%
Microsoft	32,163	71.43%	42.2%	14.28%

Table 7 reports the mean age error rate for each tool for the different threshold limits while Figure 2 highlights how the difference between the predicted age and self described age ( $\hat{a} - a$ ) varies across age groups.

We find that all the tools perform fairly poorly on age prediction for a 5 year age bin, with error rates ranging from 71% to 84%. The error rates continue to remain high even when we significantly expand the size of the acceptable age range from 5 years to 21 years. FaceX performs the worst in this scenario with an error rate of 42.19%. Further, 2 highlights that tools tend to err in favour of predicting lower age ranges, which can make age prediction harder for older population groups.



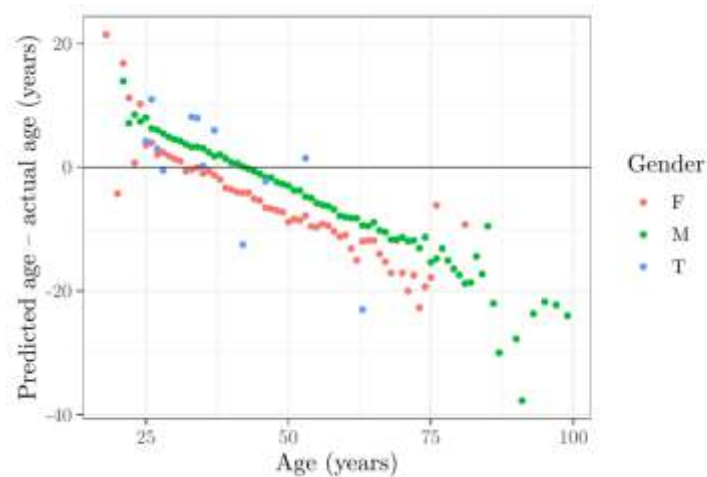


Figure 2: Difference between predicted and self described age across gender groups

Even though age prediction is generally recognised to be a hard to solve problem (Jung et al., 2018), suggested application areas of age estimation are not rare. For instance, age estimation may be used to monitor the use of age-appropriate online content (Weiss, 2021) or in recruitment processes (Angulu et al., 2018). These processes may also form the basis for automatic filtering of results for law enforcement purposes, such as identification of a suspect or a missing person based on their gender or age. Misclassification in such cases could result in incorrect identification, search or tracking, causing significant harm to the affected individual. The poorer performance of the tools in the classification of vulnerable groups like the elderly only exacerbates these concerns.

## 6 Conclusion

In this paper we audited the performance of four commercial facial processing tools, Amazon’s Rekognition, Microsoft Azure’s Face, Face++ and FaceX, on a novel India faces dataset gathered from the ECI’s election candidates portal. The audit focused on the functions of face detection, gender classification and age prediction and did not include face recognition. To the best of our knowledge, this is the first study of this nature that carries out an empirical evaluation of the performance of facial processing tools on Indian faces.

Our findings point to significant variations in the performance of the tools. There were also some clear group-specific trends observed across all tools. For face detection, Microsoft reported the highest detection error rate of 3.17%, implying an inability to correctly detect over a thousand faces in the dataset. This figure was similarly high for the Indian company, FaceX. Non-detection means that these images cannot be subjected to any further types of facial processing. While this may be a favourable outcome in certain cases, such as detection by surveillance systems, it can also lead to exclusions and hardships in the context of mandatory or hard to avoid facial processing systems.

All the tools reported higher errors in the gender classification of females compared to males and a complete inability to correctly classify persons belonging to the third gender. The female classification error was the highest in the case of China-based Face++. It had an error rate of almost 15% for Indian female faces. Besides being significantly higher than Face++’s error rate for Indian males, which was almost zero, this error rate is also much higher than that the classification errors found for females of other nationalities in a previous study. An audit conducted in 2018 reported an error rate of 2.5% for

---

Face++ using an image dataset of African and European faces (Raji and Buolamwini, 2019). This points to the lack of transferability of findings from audits done using datasets of non-Indian faces into the Indian context. It also illustrates the tendency of commercial firms to introduce band-aid measures designed to address specific issues that may be highlighted by researchers and civil society.

Age prediction errors were also high across all tools. Despite taking into account an acceptable error margin of plus or minus 10 years from a person's actual age, the tools displayed age prediction failures in the range of 14.3% to 42.2%. There was also a clear trend of predicting lower age ranges – as one grows older the tools are more likely to predict a younger age.

Future work on this subject can extend in at least two directions. The first would be to look at how facial processing interacts with other axes of diversity in the Indian society, such as caste, tribe, religion, and skin tones, and the intersection of multiple factors. The second would be to expand the scope of review beyond face detection and analysis to also include face recognition systems that carry out verification and identification functions. However, the ability to do so will have to be shaped by the limitations of available datasets and important ethical considerations.

Finally, we would like to emphasise that the accuracy of facial processing is only one among several necessary elements of a fair and accountable system. In fact, accuracy could itself be a double-edged sword, with higher accuracy leading to more rampant adoption of the technology and more exploitative use cases. Therefore, along with the need for more context-specific research directed at auditing the efficacy of facial processing systems, we must also persist with the broader agenda of questioning the validity and adoption of such systems, particularly in light of India's legal and institutional realities.

---

## References

Aneja, Urvashi and Angelina Chamuah (Jan. 2020). *We Need to Ban Facial Recognition Altogether, Not Just Regulate Its Use*. url: <https://thewire.in/tech/facial-recognition-delhi-police>.

Angulu, Raphael, Jules R. Tapamo and Aderemi O. Adewumi (June 2018). 'Age estimation via face images: a survey'. In: *EURASIP Journal on Image and Video Processing* 2018.1. doi: 10.1186/s13640-018-0278-6. url: <https://doi.org/10.1186/s13640-018-0278-6>.

De-Arteaga, Maria et al. (2019). 'Bias in bios: A case study of semantic representation bias in a high stakes setting'. In: *proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 120–128.

Barik, Soumyarendra (Apr. 2021). *Facial authentication for COVID vaccine a proof of concept, won't be mandatory: RS Sharma*. url: <https://entrackr.com/2021/04/exclusive-facial-authentication-for-covid-vaccine-a-proof-ofconcept-wont-be-mandatory-rs-sharma/>.

Basu, Arindrajit and Siddharth Sonkar (Jan. 2019). *Automated Facial Recognition Systems and the Mosaic Theory of Privacy: The Way Forward*. url: <https://aipolicyexchange.org/2019/12/30/automated-facial-recognition-systems-and-the-mosaic-theory-of-privacy-the-way-forward/>.

Bhagavatula, Suresh et al. (May 2019). *Cultural Diversity in Corporate Boards and Firm Outcomes*. url: [https://repository.iimb.ac.in/bitstream/2074/10151/1/WP\\_IIMB\\_599.pdf](https://repository.iimb.ac.in/bitstream/2074/10151/1/WP_IIMB_599.pdf).

Bhandari, Vrinda (2021). 'Facial Recognition: Why We Should Worry the Use of Big Tech for Law Enforcement'. In: *The Future of Democracy in the Shadow of Big and Emerging Tech*. Ed. by Kritika Bhardwaj, Sangh Rakshita and Shrutanjaya Bhardwaj. National Law University Delhi Press, pp. 97–113. url: <https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/the-future-of-democracy-in-the-shadow-of-big-and-emerging-tech-ccg-248.pdf>.

Buchanan, Elizabeth A. and Michael Zimmer (Jan. 2021). *Internet Research Ethics*. url: <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>.

Buolamwini, Joy and Timnit Gebru (Feb. 2018). 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification'. In: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. Ed. by Sorelle A. Friedler and Christo Wilson. Vol. 81. Proceedings of Machine Learning Research. New York, NY, USA: PMLR, pp. 77–91. url: <http://proceedings.mlr.press/v81/buolamwini18a.html>.

Chakrabarty, Amitava (Nov. 2020). *Covid-19: How this timely decision of RBI helped new investors enter financial markets*. url: <https://www.financialexpress.com/money/covid-19-how-this-timely-decision-of-rbi-helped-newinvestors-enter-financial-markets/2135609/>.

Chandna, Himani (Apr. 2021). *Modi govt now plans a 'touchless' vaccination process, with Aadhaar-based facial recognition*. url: <https://theprint.in/health/modigovt-now-plans-a-touchless-vaccination-process-with-aadhaar-basedfacial-recognition/634719/>.

Dixon, Lucas et al. (2018). 'Measuring and mitigating unintended bias in text classification'. In: *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 67–73.

ECI (Feb. 2019). Affidavit Portal. url: <https://eci.gov.in/it-applications/web-applications/affidavit-portal-r51/>.

– (Feb. 2020a). *Candidate Nomination Application*. url: <https://eci.gov.in/it-applications/web-applications/candidate-nomination-applicationr61/>.

---

– (Oct. 2020b). *Optional facility of online data entry of personal details by candidates in Nomination Form and Affidavit (in Form 26 appended to the CE Rules, 1961)*. url: <https://eci.gov.in/files/file/12313-optional-facility-of-online-data-entry-of-personal-details-by-candidates-in-nomination-form-and-affidavit-in-form-26-appended-to-the-ce-rules-1961-guidelines-all-recognized-national-state-parties-reg/>.

Fisman, Raymond, Daniel Paravisini and Vikrant Vig (Feb. 2017). ‘Cultural Proximity and Loan Outcomes’. In: *American Economic Review* 107.2, pp. 457–92. doi: 10.1257/aer.20120942. url: <https://www.aeaweb.org/articles?id=10.1257/aer.20120942>.

GoI (2010). *Face Image Data Standard for e-Governance applications in India*. Department of Information Technology, Ministry of Communications & Information Technology, Government of India.

Greig, Jonathan (May 2021). *One year after Amazon, Microsoft and IBM ended facial recognition sales to police, smaller players fill void*. url: <https://www.zdnet.com/article/one-year-after-amazon-microsoft-and-ibm-ended-facial-recognition-sales-to-police-smaller-players-fill-void/>.

Grother, Patrick, Mei Ngan and Kayee Hanaoka (Dec. 2019). *Face Recognition Vendor Test (FRVT) - Part 3 Demographic Effects*. url: <https://doi.org/10.6028/NIST.IR.8280>.

Gupta, Raj Kuwar et al. (2019). ‘IIITM Face: A Database for Facial Attribute Detection in Constrained and Simulated Unconstrained Environments’. In: *CoRR* abs/1910.01219. arXiv: 1910.01219. url: <http://arxiv.org/abs/1910.01219>.

Happy, S L et al. (2017). ‘The Indian Spontaneous Expression Database for Emotion Recognition’. In: *IEEE Transactions on Affective Computing* 8.1, pp. 131–142. doi: 10.1109/TAFFC.2015.2498174.

Hoffmann, Anna Lauren (2019). ‘Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse’. In: *Information, Communication & Society* 22.7. url: <https://doi.org/10.1080/1369118X.2019.1573912>.

Jain, Anushka (July 2020). *IFF’s Legal Notice to the NCRB on the Revised RFP for the National Automated Facial Recognition System*. url: <https://panoptic.in/case-study/iffs-legal-notice-to-the-ncrb-on-the-revised-rfp-for-the-national-automated-facial-recognition-system>.

Jain, Anushka et al. (2020). *Panoptic*. url: <https://panoptic.in/>.

Joshi, Divij (2020). *Database of Automated Decision Making Systems*. url: <https://ai-observatory.in/database>.

Jung, Soon-Gyo et al. (2018). ‘Assessing the Accuracy of Four Popular Face Recognition Tools for Inferring Gender, Age, and Race’. In: url: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/viewFile/17839/17066>.

Kalluri, Pratyusha (2020). ‘Don’t ask if artificial intelligence is good or fair, ask how it shifts power’. In: *Nature* 583.169. url: [https://www.nature.com/articles/d41586-020-02003-2?utm\\_source=twt\\_nnc&utm\\_medium=social&utm\\_campaign=naturenews](https://www.nature.com/articles/d41586-020-02003-2?utm_source=twt_nnc&utm_medium=social&utm_campaign=naturenews).

Katti, Harish and S. P. Arun (July 2019). ‘Are you from North or South India? A hard face-classification task reveals systematic representational differences between humans and machines’. In: *Journal of Vision* 19.7, pp. 1–1. issn: 1534-7362. doi: 10.1167/19.7.1. eprint: [https://arvojournals.org/arvo/content\\_public/journal/jov/938093/i1534-7362-19-7-1.pdf](https://arvojournals.org/arvo/content_public/journal/jov/938093/i1534-7362-19-7-1.pdf). url: <https://doi.org/10.1167/19.7.1>.

Keyes, OS (2018). ‘The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition’. In: url: <https://dl.acm.org/citation.cfm?id=3274357>.

- 
- Khalil, Ashraf et al. (June 2020). 'Investigating Bias in Facial Analysis Systems: A Systematic Review'. In: *IEEE Access* 8. url: <https://ieeexplore.ieee.org/document/9130131>.
- Khan, Zaid and Yun Fu (Feb. 2021). *One Label, One Billion Faces: Usage and Consistency of Racial Categories in Computer Vision*. url: <https://arxiv.org/abs/2102.02320>.
- Khera, Reetika (Apr. 2019). 'Aadhaar Failures: A Tragedy of Errors'. In: 54 (14). url: <https://www.epw.in/engage/article/aadhaar-failures-food-serviceswelfare>.
- Klare, B. F. et al. (2012). 'Face Recognition Performance: Role of Demographic Information'. In: *IEEE Transactions on Information Forensics and Security* 7.6, pp. 1789–1801. doi: 10.1109/TIFS.2012.2214212.
- Kovacs, Anja (May 2020). *When our bodies become data, where does that leave us?* url: <https://deepdives.in/when-our-bodies-become-data-where-doesthat-leave-us-906674f6a969>.
- Lakshmi, Anjana et al. (Feb. 2021). 'The India Face Set: International and Cultural Boundaries Impact Face Impressions and Perceptions of Category Membership'. In: *Frontiers in Psychology* 12. url: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7905305/>.
- Lalwani, Vijayta (July 2019). *Facial recognition: As airports in India start using the technology, how will it be regulated?* url: <https://scroll.in/article/929851/facial-recognition-as-airports-in-india-start-using-the-technology-how-will-it-be-regulated>.
- Leufer, Daniel (Apr. 2021). *Computers are binary, people are not: how AI systems undermine LGBTQ identity*. url: <https://www.accessnow.org/how-ai-systems-undermine-lgbtq-identity/>.
- Manikandan, Ashwin and Saloni Shukla (Oct. 2020). *Facial recognition, iris scans may be used for welfare scheme payouts*. url: <https://economictimes.indiatimes.com/industry/banking/finance/banking/facial-recognition-iris-scans-may-be-used-for-welfare-scheme-payouts/printarticle/77755102.cms>.
- Marda, Vidushi (Nov. 2019a). *Indian govt's approach to facial recognition is flawed & driven by faulty assumptions*. url: <https://theprint.in/opinion/indiangovt-approach-to-facial-recognition-flawed-driven-by-faulty-assumptions/327036/>.
- (2019b). *Introduction*. url: [https://giswatch.org/sites/default/files/gisw2019\\_web\\_intro\\_0.pdf](https://giswatch.org/sites/default/files/gisw2019_web_intro_0.pdf).
- Marda, Vidushi and Shazeda Ahmed (2021). *Emotional Entanglement: China's emotion recognition market and its implications for human rights*. url: <https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf>
- Marda, Vidushi and Shivangi Narayan (2020). 'Data in New Delhi's Predictive Policing System'. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. FAT\* '20. Barcelona, Spain: Association for Computing Machinery, pp. 317–324. isbn: 9781450369367. doi: 10.1145/3351095.3372865. url: <https://doi.org/10.1145/3351095.3372865>.
- Moreau, Yves (Dec. 2019). 'Crack down on genomic surveillance'. In: *Nature* 576, pp. 36–38. doi: <https://doi.org/10.1038/d41586-019-03687-x>.
- Narayanan, Dhanya Lakshmi et al. (Dec. 2019). 'Computer-aided Facial Analysis in Diagnosing Dysmorphic Syndromes in Indian Children'. In: *Indian Pediatrics* 56.12, pp. 1017–1019. doi: 10.1007/s13312-019-1682-4. url: <https://doi.org/10.1007/s13312-019-1682-4>.
- National Legal Services Authority vs Union of India (Apr. 2014). *Writ Petition (Civil) No. 400 of 2012*. url: <https://indiankanoon.org/doc/193543132//>.

---

Noorden, Richard Van (Nov. 2020). 'The ethical questions that haunt facial-recognition research'. In: *Nature* 587, pp. 354–358. doi:

O'Toole, Alice (2011). 'Face Recognition by Humans and Machines'. In: *Handbook of Face Recognition*. Ed. by Stan Z. Li and Anil K. Jain. Second.

Parsheera, Smriti (2018). 'A Gendered Perspective on Artificial Intelligence'. In: *2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K)*. IEEE Xplore, pp. 1–7. doi: 10.23919/ITU-WT.2018.8597618. url: <https://ieeexplore.ieee.org/document/8597618>.

– (2019). *Adoption and regulation of facial recognition technologies in India: Why and why not*. url: <https://datagovernance.org/report/adoption-%20andregulation-of-facial-recognition-technologies-in-india>.

– (2021). 'Codes and Coalitions: Path to Global Governance of Artificial Intelligence?' In: *Digital Debates: CyFy Journal* 8, pp. 55–65. url: [https://www.orfonline.org/wp-content/uploads/2021/10/Digital-Debates\\_\\_CyFy2021.pdf](https://www.orfonline.org/wp-content/uploads/2021/10/Digital-Debates__CyFy2021.pdf).

Prates, Marcelo O.R., Pedro H Avelar and Luis C Lamb (2019). 'Assessing gender bias in machine translation: a case study with google translate'. In: *Neural Computing and Applications*, pp. 1–19.

PTI (Aug. 2018). *Delhi police facial recognition software has only 2 per cent accuracy: HC told*. url: [https://www.business-standard.com/article/pti-stories/delhi-police-facial-recognition-software-has-only-2-per-cent-accuracy-hc-told-118082301289\\_1.html](https://www.business-standard.com/article/pti-stories/delhi-police-facial-recognition-software-has-only-2-per-cent-accuracy-hc-told-118082301289_1.html).

PTI (Aug. 2019). *Police facial recognition software glitchy: Centre*. url: <https://www.thehindu.com/news/cities/Delhi/police-facial->

Raji, Inioluwa Deborah and Joy Buolamwini (2019). 'Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products'. In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. AIES '19. Honolulu, HI, USA: Association for Computing Machinery, pp. 429–435. isbn: 9781450363242. doi: 10.1145/3306618.3314244. url: <https://doi.org/10.1145/3306618.3314244>.

Raji, Inioluwa Deborah, Timnit Gebru et al. (2020). 'Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing'. In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. AIES '20. New York, NY, USA: Association for Computing Machinery, pp. 145–151. isbn: 9781450371100. doi: 10.1145/3375627.3375820. url: <https://doi.org/10.1145/3375627.3375820>.

Ravn, Signe, Ashley Barnwell and Barbara Barbosa Neves (2020). 'What Is "Publicly Available Data"? Exploring Blurred Public–Private Boundaries and Ethical Practices Through a Case Study on Instagram'. In: *Journal of Empirical Research on Human Research Ethics* 15.1-2, pp. 40–45. doi: <https://doi.org/10.1177/1556264619850736>.

Romine, Charles H. (Feb. 2020). *Facial Recognition Technology (FRT) - NIST testimony to Committee on Homeland Security, United States House of Representatives*. url: <https://www.nist.gov/speech-testimony/facial-recognition-technology-frt-0>.

Saha, Kankan et al. (2013). 'Creation of North-East Indian face database for human face identification'. In: *Sensors, Cameras, and Systems for Industrial and Scientific Applications XIV*. Ed. by Ralf Widenhorn and Antoine Dupret. Vol. 8659. International Society for Optics and Photonics. SPIE, pp. 253–258. doi: 10.1117/12.2002028. url: <https://doi.org/10.1117/12.2002028>.

Sambasivan, Nithya et al. (2021). 'Re-Imagining Algorithmic Fairness in India and Beyond'. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '21. Virtual Event, Canada: Association for Computing Machinery, pp. 315–328. isbn: 9781450383097. doi: 10.1145/3442188.3445896. url: <https://doi.org/10.1145/3442188.3445896>.

---

Sap, Maarten et al. (2019). 'The risk of racial bias in hate speech detection'. In: *Proceedings of the 57th annual meeting of the association for computational linguistics*, pp. 1668–1678.

Sarin, Samarth and Siba Panda (2020). 'A Convolution Neural Network for Classification of Indian Faces'. In: *International Journal of Computer Applications* 177 (24). url: <https://www.ijcaonline.org/archives/volume177/number24/sarin-2019-ijca-919693.pdf>.

Scheuerman, Morgan Klaus, Jacob M. Paul and Jed R. Brubaker (Nov. 2019). 'How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis and Image Labeling Services'. In: *Proceedings of the ACM on Human-Computer Interaction* 3. doi: <https://doi.org/10.1145/3359246>.

Setty, S. et al. (2013). 'Indian Movie Face Database: A benchmark for face recognition under wide variations'. In: *2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, pp. 1–5. doi: [10.1109/NCVPRIPG.2013.6776225](https://doi.org/10.1109/NCVPRIPG.2013.6776225).

Sharma, Reecha and M. S. Patterh (Sept. 2015). 'Article: Indian Face Age Database: A Database for Face Recognition with Age Variation'. In: *International Journal of Computer Applications* 126.5. Published by Foundation of Computer Science (FCS), NY, USA, pp. 21–27.

Singh, Shivendra and Shajulin Benedict (2020). 'Indian Semi-Acted Facial Expression (iSAFE) Dataset for Human Emotions Recognition'. In: *Advances in Signal Processing and Intelligent Recognition Systems*. Ed. by Sabu M. Thampi et al. Singapore: Springer Singapore, pp. 150–162.

Snow, Jacob (2018). *Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots*. url: <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28>.

Sweeney, Patrick (2017). 'Images of Faces Gleaned from Social Media in Social Psychological Research on Sexual Orientation'. In: *Internet Research Ethics for the Social Age: New Challenges, Cases and Contexts*. Ed. by Michael Zimmer and Katharina Kinder-Kurlanda. Peter Lang Publishing, pp. 287–293.

Vipra, Jai (Aug. 2021). *The Use of Facial Recognition Technology for Policing in Delhi: An Empirical Study of Potential Religion-Based Discrimination*. url: <https://vidhilegalpolicy.in/research/the-use-of-facial-recognition-technology-for-policing-in-delhi/>.

Vitak, Jessica, Katie Shilton and Zahra Ashktorab (2016). 'Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community'. In: *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, pp. 941–953.

Wang, Cunrui et al. (Jan. 2019). 'Facial feature discovery for ethnicity recognition'. In: *Wire's Data Mining and Knowledge Discovery* 9.1. doi: <https://doi.org/10.1002/widm.127>.

Wang, Yilun and Michal Kosinski (2018). 'Deep neural networks are more accurate than humans at detecting sexual orientation from facial images'. In: *Nature* 114.2, pp. 246–257. doi: <https://doi.org/10.1037/pspa0000098>.

Weiss, Eric (Mar. 2021). *Yoti Takes Steps to Improve Age Estimation Technology*. url: <https://findbiometrics.com/yoti-takes-steps-improve-ageestimation-technology-032901/>.

Xia, Mengzhou, Anjalie Field and YuliaT svetkov (2020). 'Demoting Racial Bias in Hate Speech Detection'. In: *arXiv preprint arXiv:2005.12246*.

Zimmer, Michael (2010). '"But the data is already public": on the ethics of research in Facebook'. In: *Ethics and Information Technology* 12, pp. 313–325. doi: <https://doi.org/10.1007/s10676-010-9227-5>.

 [datagovernance.org](https://datagovernance.org)    [dgn@idfcinstitute.org](mailto:dgn@idfcinstitute.org)

 [@datagovnetwork](https://twitter.com/datagovnetwork)    [/datagovnetwork](https://facebook.com/datagovnetwork)    [/datagovnetwork](https://youtube.com/datagovnetwork)