How do Face Datasets Inform Attractiveness?

AC 221

Jingxing Fang
Aditi Memani
Hari Raval

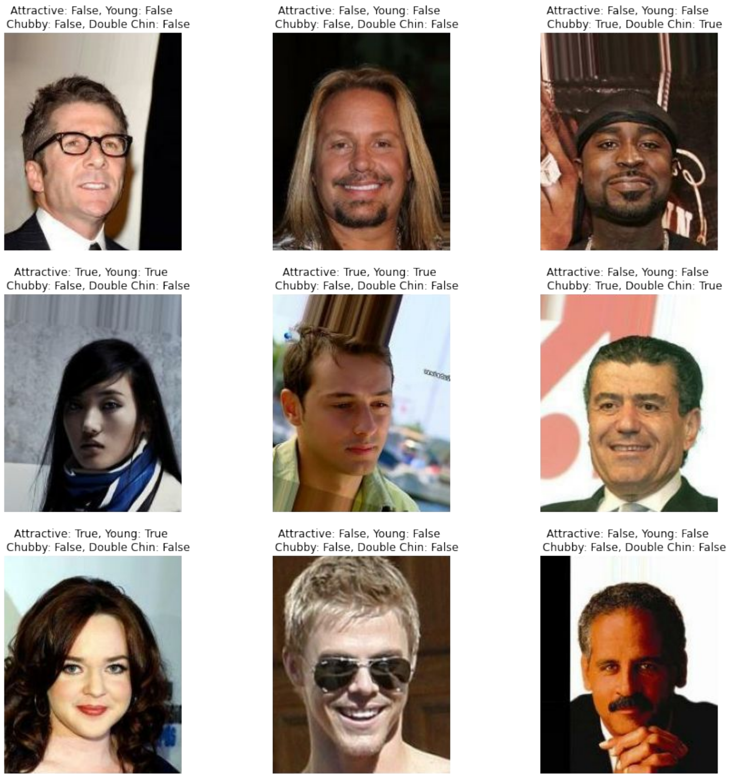
Introduction

Motivation

- Historical and representational biases exist in datasets
- CelebA dataset contains such biases
- Attributes, like attractiveness, are impossible to evaluate quantitatively
- Unfair advantage for certain subsets of the population
    - *Young* vs. *Old*
    - *Male* vs. *Female*
    - *Blonde* vs. *Not Blonde*
    - *Not Chubby* vs. *Chubby*
    - *Double Chin* vs. *No Double Chin*

CelebA Dataset

- Large-scale face attributes dataset
  - 202,599 image samples
  - 10,177 unique identities
  - Large pose variations
- Various attribute values
  - Over 40 binary attributes per image such as *hair color*, *sex*, and *age*
  - Certain subjective attributes such as *attractiveness*
- Available as a TensorFlow dataset

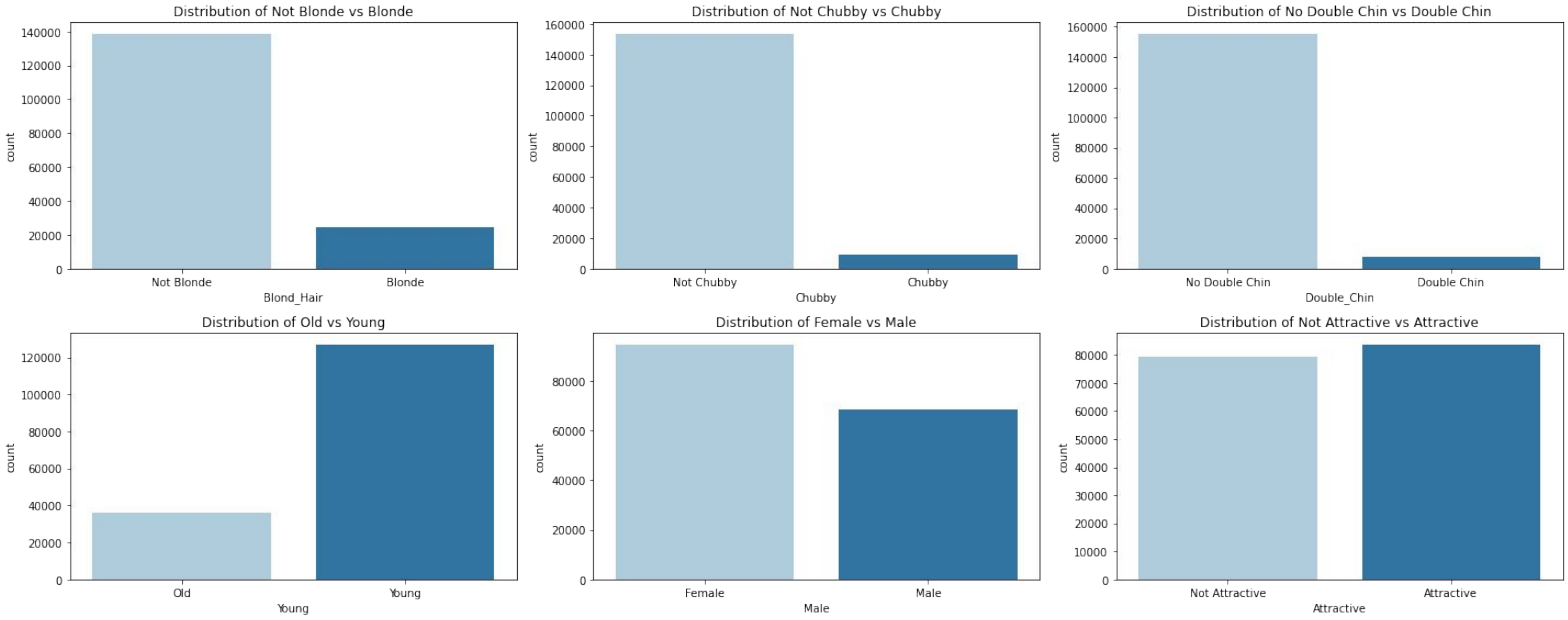


Goals

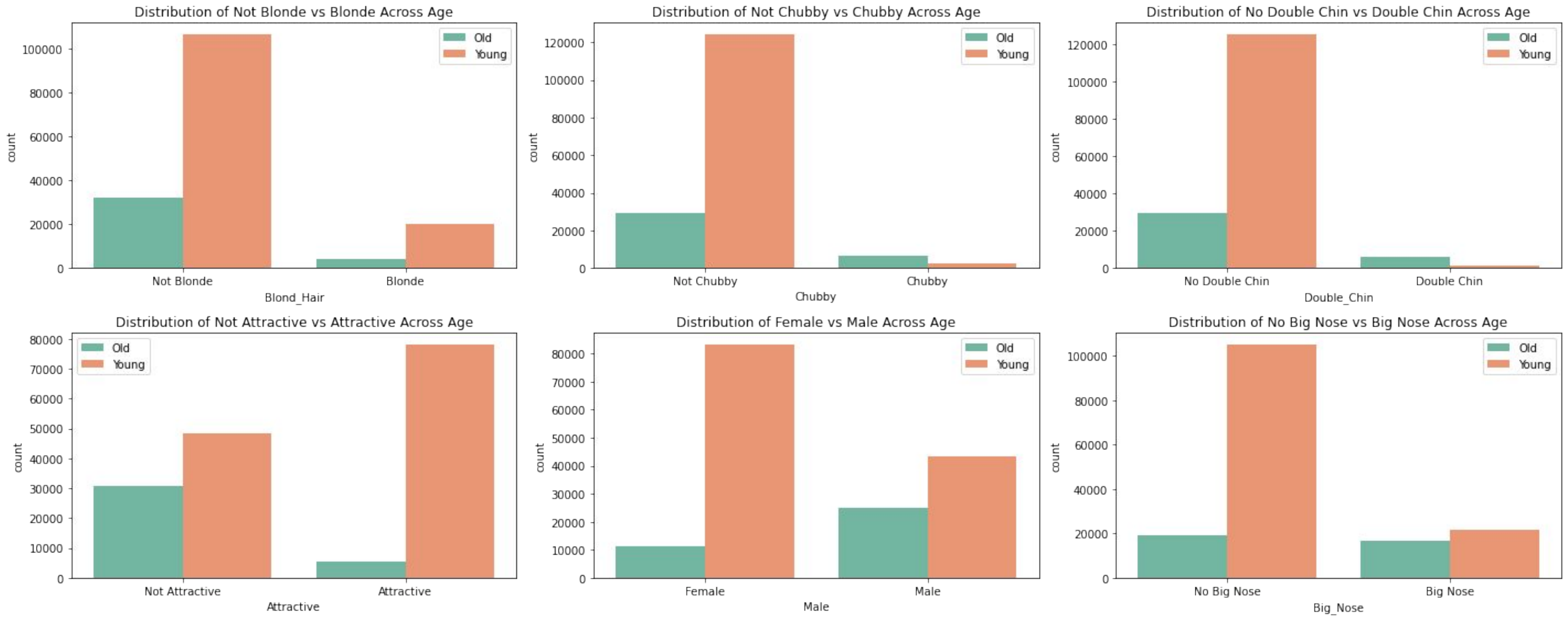
- Investigate whether biases in our CelebA training data will present themselves in the models we create
    - Build a CNN model to predict attractiveness
    - Evaluate subgroup metrics
- Examine the real-world implications of our findings
- Highlight the importance of detecting dataset biases early on
    - Adjust for such biases when creating models
    - Restart the data collection process

Exploratory Data Analysis

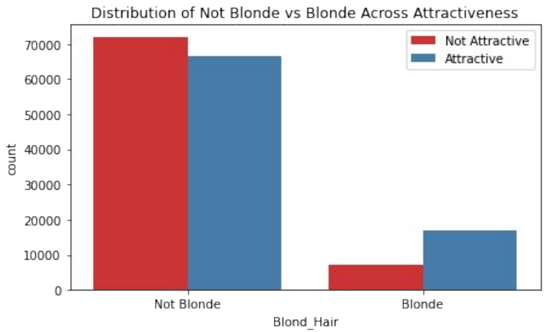
Distribution of Characteristics

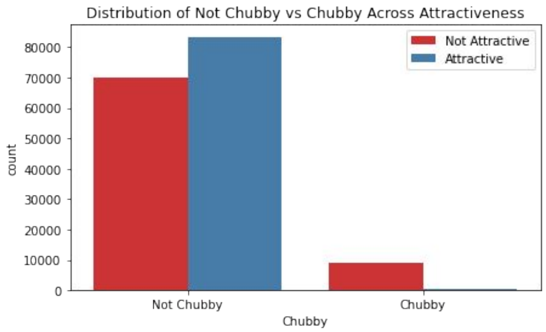


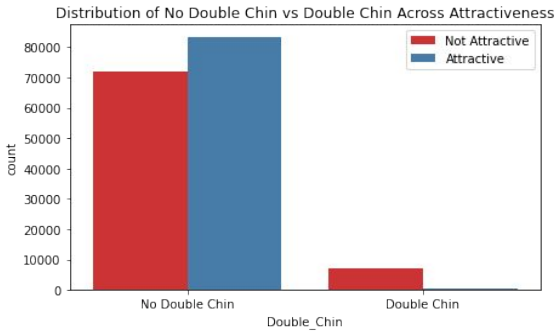
Characteristics Across Age

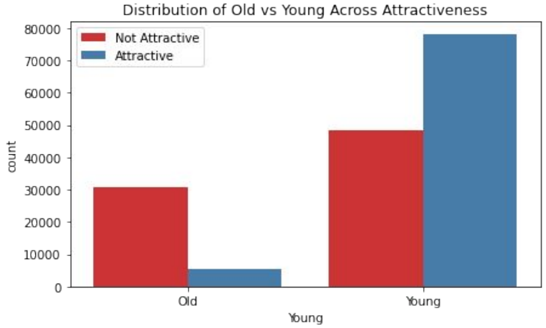


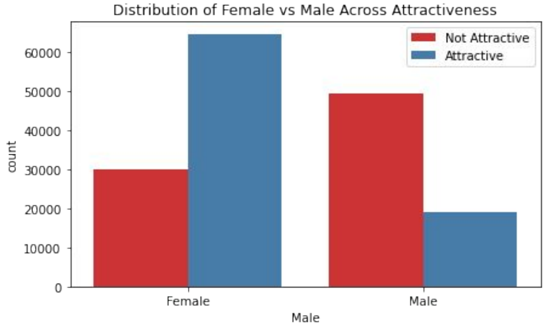
Characteristics Across Attractiveness

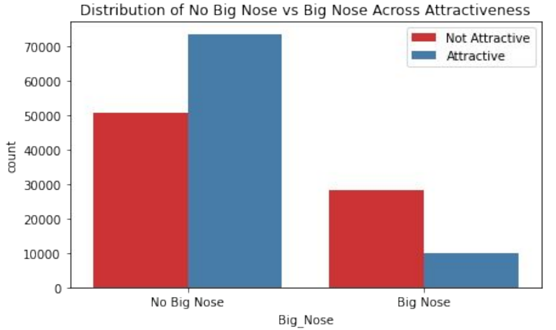




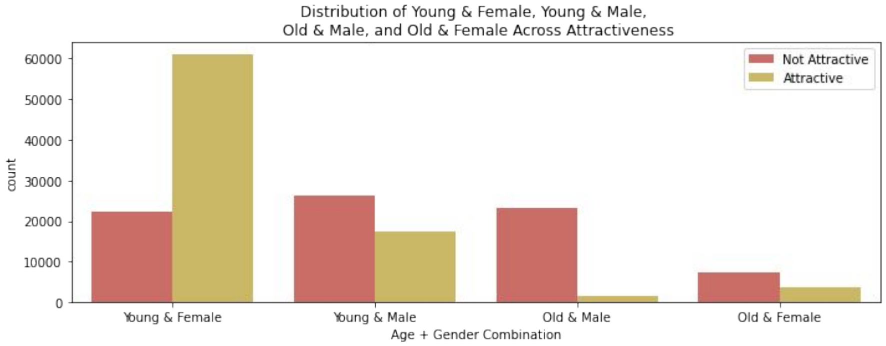


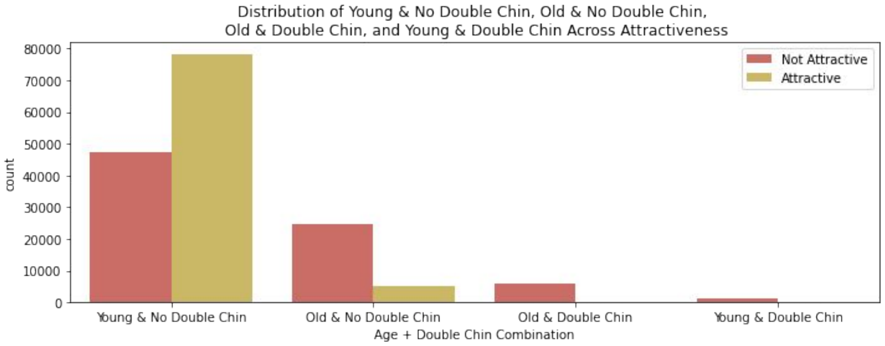


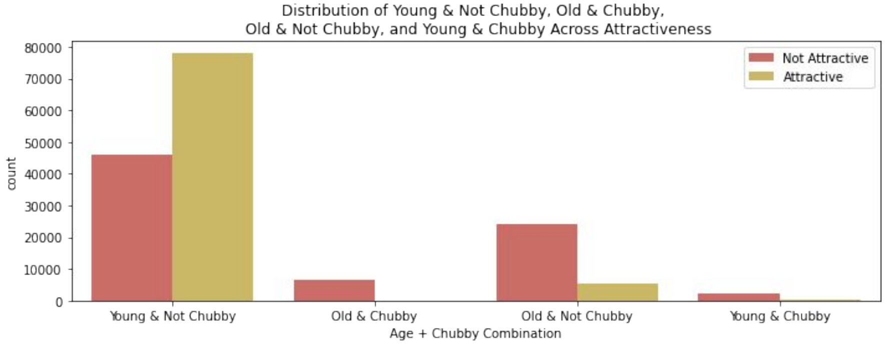


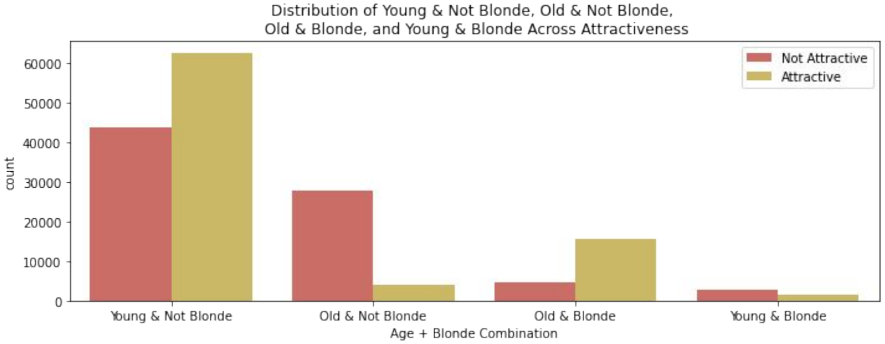


Grouped Characteristics Across Attractiveness



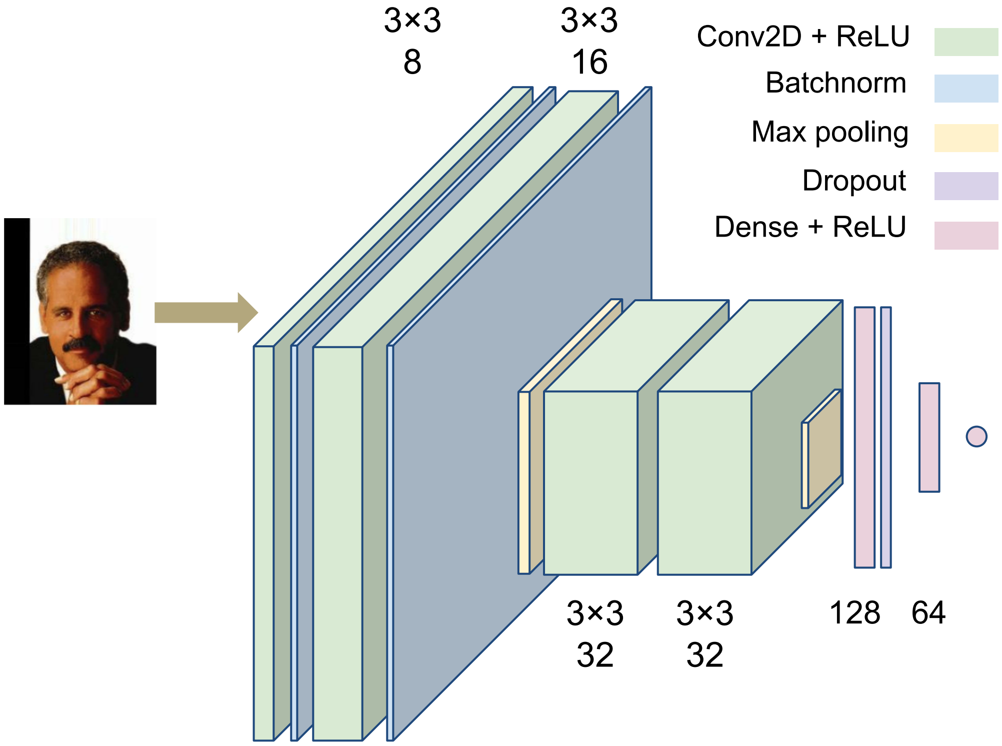






Model Development

Design and Training



- 4 Conv. layers + 2 Dense layers
- Total trainable parameters: 1,891,553
- Trained 2 epochs for ~ 8 hours
- Training accuracy: 79.79%
- Validation accuracy: 79.40%

Model Limitations

- Limited access to computational resources
- Limited number of training data
- Interpretability
- CNN’s lack of ability to understand depth
- Vulnerability to adversarial attack

Model Evaluation

Accuracy Results

Overall test dataset accuracy of 80.61%

The proportion of being predicted as attractive

V.S.

The proportion of actually being attractive (labeled attractive in the dataset)

	Predicted	Actual
Male	0.178743 <	0.248088
Female	0.695027	0.651915

	labels	Predicted	Actual
0	Old & Female	0.405010	0.322025
1	Old & Male	0.017053 <	0.056276
2	Young & Female	0.748814	0.713097
3	Young & Male	0.277859 <	0.365670

	Predicted	Actual
Blond_Hair	0.686466	0.653759
Not Blonde	0.466131 <	0.471564

Error Rates Analysis

Rates of interest:

- False Positive Rate (FPR): $$FPR = \frac{FP}{FP + TN}$$
- False Negative Rate (FNR): $$FNR = \frac{FN}{FN + TP}$$

Error Rates Analysis

	Male	Female
acc	0.823590	0.795052
fpr	0.071195	0.356322
tpr	0.504702	0.875877
fnr	0.495298	0.124123

	Blond_Hair	Not Blonde
acc	0.792105	0.808230
fpr	0.347448	0.176310
tpr	0.866015	0.790906
fnr	0.133985	0.209094

	Young & Female	Young & Male	Old & Female	Old & Male
acc	0.801471	0.748693	0.760438	0.945771
fpr	0.408232	0.128873	0.237875	0.007951
tpr	0.885842	0.536306	0.756888	0.169697
fnr	0.114158	0.463694	0.243112	0.830303

Key Findings

Error Rate Analysis - One Attribute

- Representation Biases in the training dataset were further amplified by the model
    - Significant false positive rates:
        - Young, Female, Blonde Hair, Not Chubby, and No Double Chin
        - Systematically biased towards these labels
    - Significant false negative rates:
        - Old, Male, Chubby, and Double Chin
        - Systematically biased against these labels
    - Outlier: Not Blonde

Error Rate Analysis - Two Attributes

- Disparity within Age is significant
  - Significant false positive rates:
    - Young & Female, Young & Blonde Young & Not Chubby, and Young & No Double Chin
    - Systematically biased towards these labels
  - Significant false negative rates:
    - Old & Male, Old & Not Blonde and Old & Double Chin
    - Systematically biased against these labels
  - Outlier: Young & Chubby
    - Due to Underrepresentation

Real World Implications

- Joy Buolamwini: Analysis of 3 algorithms
  - Failed to recognize dark-skinned individuals
- NIST: Federal study of 189 Algorithms
  - Confirmed concerns about training dataset biases
  - Oversampling of minority groups – used by law enforcement
  - US algorithms vs. Chinese Algorithms – Asian Individuals
- Easily mitigated by using evenly distributed datasets
- Algorithmic Accountability Act (2019)

Conclusion

Takeaways

- Dataset biases propagated through the CNN model
    - Biased towards the 'Young', 'Female', 'Blonde Hair', 'Not Chubby', and 'No Double Chin'
    - Biased against the 'Old', 'Male', 'Not Blonde', 'Chubby', 'Double Chin
- Difficult to detect such biases since they match expectations
    - Scrutinize the error rates
- Our findings match existing research conclusions
- Monitor how biases could enter at every stage of development
- Must mitigate to reduce discrimination and marginalization

Citations

[1] Hardesty, Larry. “Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems.” MIT News, February 11, 2018. Available at http://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212 (last accessed April 19, 2019). These companies were selected because they provided gender classification features in their software and the code was publicly available for testing.

[2] Singer, Natasha, and Cade Metz. “Many Facial-Recognition Systems Are Biased, Says U.S. Study.” The New York Times, 19 Dec. 2019, www.nytimes.com/2019/12/19/technology/facial-recognition-bias.html.

[3] Crumpler, William. “The Problem of Bias in Facial Recognition.” Www.csis.org, 1 May 2020, www.csis.org/blogs/technology-policy-blog/problem-bias-facial-recognition.

[4] Review, The Regulatory. “Facing Bias in Facial Recognition Technology | the Regulatory Review.” Www.theregreview.org, 20 Mar. 2021, www.theregreview.org/2021/03/20/saturday-seminar-facing-bias-in-facial-recognition-technology/.

[5] Perkowitz, Sidney. “The Bias in the Machine: Facial Recognition Technology and Racial Disparities.” MIT Case Studies in Social and Ethical Responsibilities of Computing, 5 Feb. 2021, 10.21428/2c646de5.62272586

[6] https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

Thank You

Questions?