THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Auditing Biases in Computer Vision: How CelebA Encodes and Reproduces Gender Stereotypes in Appearance

A dissertation submitted to the London School of Economics and Political Science for the degree of Master of Science in the Department of Methodology

Applied Social Data Science

Candidate Number: 45774

Supervisor: Dr. Yuanmo He

Word count: 9869

Abstract

Large-scale facial datasets like CelebA are widely used in computer vision, yet their embedded cultural biases remain underexplored. This study audits how gendered double standards of ageing and beauty, rooted in media representation, are encoded in CelebA and reproduced in model behaviour. Using hierarchical clustering of 202,599 images (Study 1a), seven latent trait bundles emerge, reflecting societal archetypes: *performative femininity* (youth, makeup, adornment) and *professional masculinity* (aging, facial hair, formalwear). A cluster-based analysis (Study 1b) shows that female faces, though more often rated attractive overall, incur disproportionately steep penalties when assigned to ageing or masculine-coded clusters, while pale skin provides an uplift. Training XGBoost classifiers with SHAP analysis (Study 2) reveals gender-specific effects, such as adiposity reducing attractiveness only for females. Subgroup evaluation with Grad-CAM (Study 3) finds that predictions for females concentrate on mid-face cues, whereas those for males shift toward hair and clothing. Together, the findings underscore the need to address representational harms alongside performance disparities in fairness research.

Keywords: representation bias; computer vision; gender bias; algorithm audit; explainable AI