Age-based Gender Inequality in over One Million Images from Google, Wikipedia, and IMDb

Keywords: Gender Stereotypes; Inequality; Search Engines; Categorization; Image Analytics

Extended Abstract

Women face disproportionate pressure to appear young, constituting a pervasive bias that has been linked to workplace discrimination and detrimental effects on the mental health of women[1]. Yet, despite its prevalence and impacts, age-related gender inequality remains under-appreciated and unexplored in audits of gender bias online. Here, we identify associations between age and gender in over one million images from Google, Wikipedia, and IMDb, associated with over 3,000 common social categories, including occupations (e.g., “doctor”) as well as generic social roles (e.g., “friend”) and lifestyles (e.g., “vegan”). Across categories and platforms, women were associated with significantly younger ages than men, according to classification judgments from human annotators and a deep learning algorithm, as well as objective measures of the actual age and gender of the people depicted, as derived from ground-truth datasets containing images with time-stamped metadata. These findings replicate when analyzing Google search results from six distinct cities around the world, including New York, Bangalore, Amsterdam, Frankfurt, Toronto, and Singapore. Our study reveals a crucial pathway for the large-scale diffusion of gender bias, since images from these platforms are frequently viewed, downloaded, and circulated, and have the capacity to reinforce biases in people’s beliefs[2,3].

Methods: To examine the correlation between gender and age in Google Images, we collected the top 100 images associated with each of the 3,435 social categories contained within Wordnet, a lexical ontology that maps the taxonomic structure of the English language. These categories include occupations (such as “physicist”) as well as generic social roles (such as “colleague”). This yielded 641,122 unique images containing faces from Google. Searches were run from 10 distinct data servers in New York City. To collect images from Wikipedia, we identified the images associated with each social category in the 2021 “Wikipedia-based Image Text Dataset” (WIT)[4]. WIT maps all images over Wikipedia to textual descriptions based on the title, content, and metadata of the active Wikipedia articles in which they appear. WIT contained images associated with 1,523 social categories from Wordnet across all English Wikipedia articles. This yielded 15,609 unique Wikipedia images.

We hired 6,392 human annotators from Amazon’s Mechanical Turk to classify the gender and age of the faces in these images. Each face was classified by three unique annotators, so that the gender of each face (“Male” or “Female”) could be identified based on the majority gender classification across three coders (we also gave coders the option of labeling the gender of faces as “Non-binary”, but this option was only chosen in less than 2% of cases, so we excluded this data from our main analyses). In terms of age, each face was classified as belonging to one of the following age bins: (1) 0-11, (2) 12-17, (3) 18-24, (4) 25-34, (5) 35-54, (6) 55-74, (7) 75+ years old. We identified the age of each face by taking the average of the ordinal age bin judgments across the three coders.

We extend our findings by examining age-related gender bias in two large, independently collected corpora of online images for which the ground-truth gender and age of the faces is objectively inferred without the use of human annotators or machine learning. First, we analyzed the 2018 IMDb-Wiki dataset[5], which consists of over half a million images of celebrities from IMDb and Wikipedia, focusing on those depicted in the top 100,000 most visited IMDb pages. Second, we analyzed the 2014 Cross-Age Celebrity Dataset (CACD)[6], which consists of 163,446 images collected from the Google search engine depicting 2,000 celebrities, comprising the top 50 most popular celebrities each year from 1951 to 1990. Each of these datasets identifies the exact age of the celebrities at the time they were depicted in each photo by determining the date-of-birth and gender of each celebrity on their public IMDb and Wikipedia pages, and then by comparing this information to the time-stamped date of when each photo of each celebrity was taken.

Main results: Women were identified as significantly younger than men in Google Images (Figure 1*A*; $p<.0001$, t=-73.87, CI=[-0.37,-0.35], 159,867 images, 3,435 categories, Student's t-test, Two-sided). We replicated our methodology while searching for the male and female version of each category in Google Images (e.g., by searching for the top 100 Google images associated "male athlete" and "female athlete" separately), yielding 483,353 images. Again, younger women were over-represented in these explicitly-gendered Google Image search results (Figure 1*B*; $p<.0001$, t=-36.57, CI=[-0.3,-0.27], paired Student's t-test, Two-sided, results are paired at the category level). We replicate this finding using images from Wikipedia, where again women were coded as significantly younger than men across 1,523 social categories (Figure 1*C*; $p<.0001$, t=-39.46, CI=[-0.76,-0.69], 15,609 images).

We also replicate these findings across IPs. Women were identified as significantly younger than men in Google images from searches in New York (Figure *2A*, $p<.001$, t=-38.4, CI=[-0.46,-0.42], N=12,860 images), Singapore (Figure *2B*, $p<.001$, t=-24.9, CI=[-0.61,-0.52], N=9,461 images), Frankfurt (Figure *2C*, $p<.001$, t=-23.3, CI=[-0.54,-0.45], N=10,820 images), Bangalore (Figure *2D*, $p<.001$, t=-25.2, CI=[-0.62,-0.53], N=9,170 images), Toronto (Figure *2E*, $p<.001$, t=-25.5, CI=[-0.57,-0.49], N=11,232 images), and Amsterdam (Figure *2F*, $p<.001$, t=-19.7, CI=[-0.49,-0.40], N=10,052 images), (Student's t-test, Two-sided).

Lastly, we show that female celebrities are significantly younger than male celebrities in nearly half a million online images from IMDb (Figure 3A, average age difference of 6.4 years, $p<.00001$, t=-169.09, CI=[-6.43,-6.28], 451,570 images) and Wikipedia (Figure 3*B*, average age difference of 3.4 years, $p<.00001$, t=-20.24, CI=[-3.6,-3.0], 57,932 images), (Student's t-test, Two-sided). Similarly, using the Cross-Age Celebrity dataset of Google images[6], we again find that female celebrities are significantly younger than male celebrities online (Figure 3*C*, average age difference of 5.4 years, $p<.00001$, t=-90.92, CI=[-5.47,-5.24], 149,889 images, Student's t-test, Two-sided).

References

1. Gendered ageism: A double jeopardy for women in organizations. in *Gender, Culture and Organizational Change* (eds. Itzen, C. & Newman, J.) 93–103 (Routledge, 1995).
2. Metaxa, D., Gan, M. A., Goh, S., Hancock, J. & Landay, J. A. An Image of Society: Gender and Racial Representation and Impact in Image Search Results for Occupations. *Proc. ACM Hum.-Comput. Interact.* **5**, 26:1-26:23 (2021).
3. Vlasceanu, M. & Amodio, D. M. Propagation of societal gender inequality by internet search algorithms. *PNAS* **119**, e2204529119 (2022).
4. Srinivasan, K., et al. WIT: Wikipedia-based Image Text Dataset for Multimodal Multilingual Machine Learning. in *P of the 44th Int'l ACM SIGIR Conference on Research and Development in Information Retrieval* 2443–2449.
5. Rothe, R., Timofte, R. & Van Gool, L. Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks. *Int J Comput Vis* **126**, 144–157 (2018).
6. Chen, B.-C., Chen, C.-S. & Hsu, W. H. Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval. in *Computer Vision – ECCV 2014* (eds. Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T.) 768–783 (Springer International Publishing, 2014). doi:10.1007/978-3-319-10599-4_49.

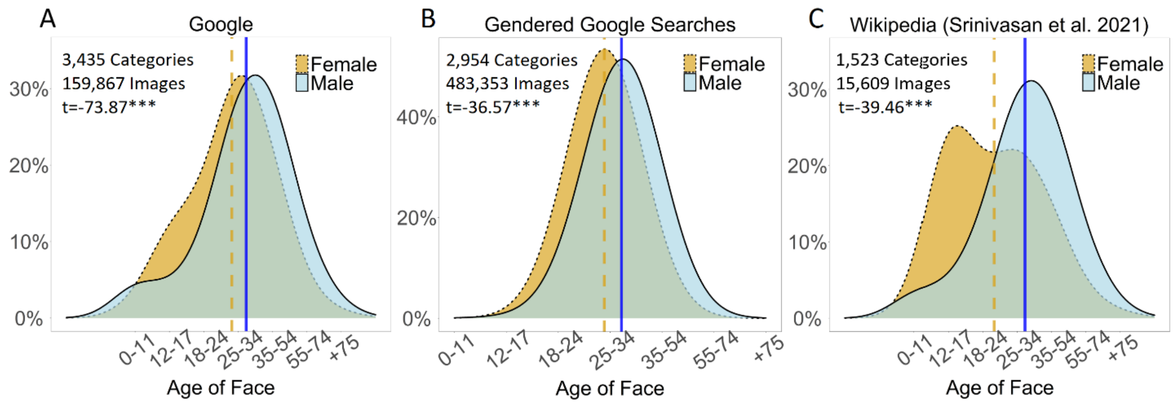

Figure 1: Human judgments of age and gender for (A) the top 100 Google images associated with 3,435 social categories (159,867 images), (B) the top 100 Google images retrieved using gendered searches (e.g., by searching "female athlete" or "male athlete") shown for all non-gendered categories in WordNet (*N*=2,954; 483, 353 images), and (C) 1,523 categories in Wikipedia (from the Srinivasan et al. 2021 Wikipedia-based Image Text dataset; 15,609 images). Dashed gold (solid blue) lines indicate the average age for female (male) faces according to each dataset. ***, p<.000001. These results were from Google searches run from New York, USA.

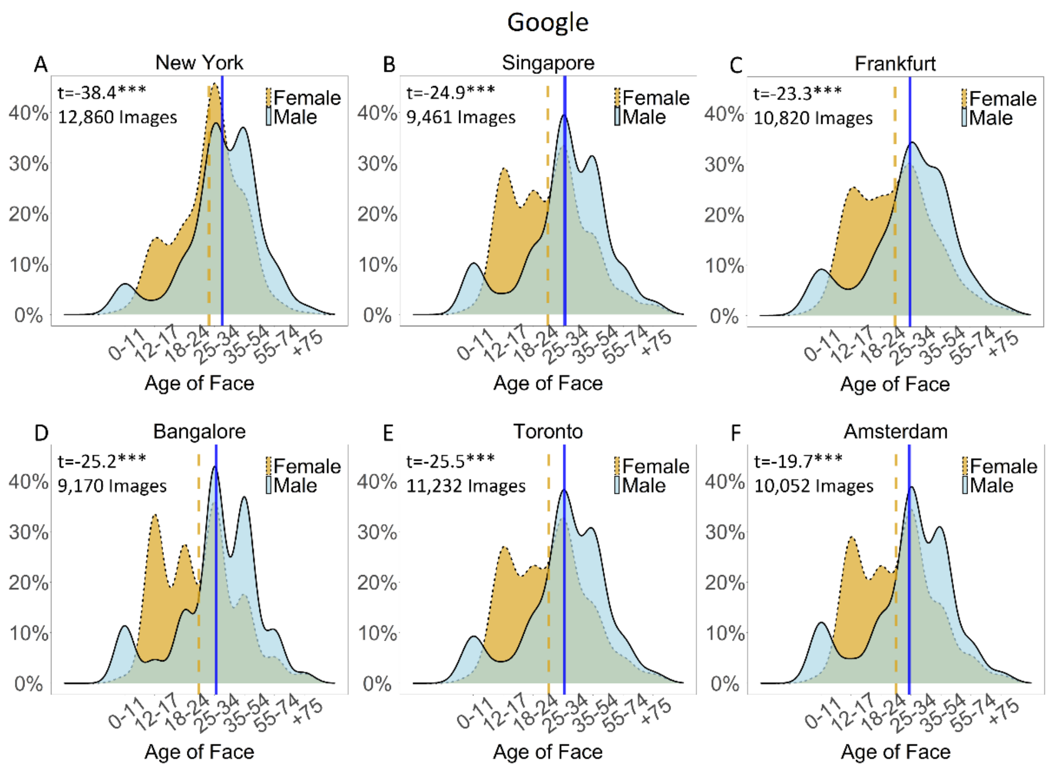

Figure 2: Human judgments of age and gender are shown for the same 300 social categories in Google Images collected while searching from an IP address based in (A) New York (12,860 images), (B) Singapore (9,461 images), (C) Frankfurt (10,820 images), (D) Bangalore (9,170 images), (E) Toronto (11,232 images), and (F) Amsterdam (10,052 images). Dashed gold (solid blue) lines indicate the average age for female (male) faces according to each dataset. ***, p<.00001.

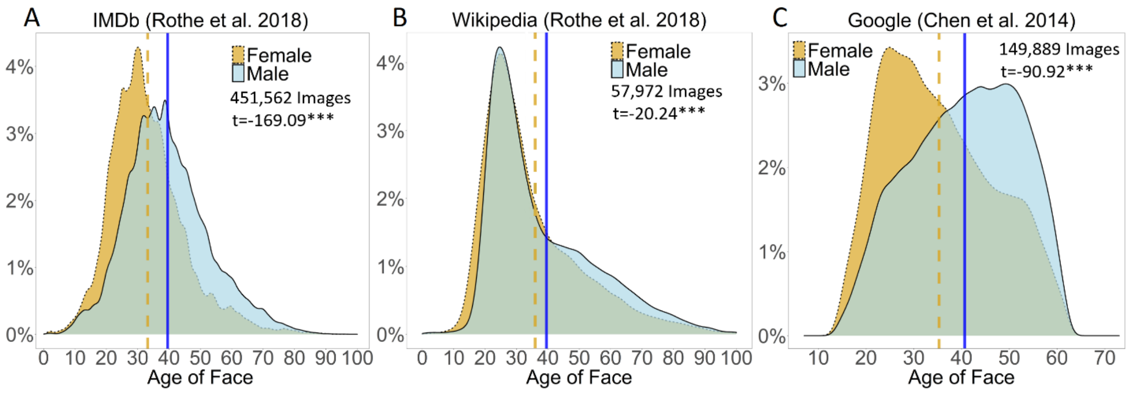

Figure 3: Patterns of gendered ageism in online images using canonical ground-truth datasets. The correlation between age and gender is shown for (A) the top 100,000 most popular pages on IMDb (451,570 images), (B) the biographical Wikipedia pages describing these same celebrities, according to the IMDb-Wiki dataset (57,932 images), and (C) the top 50 most popular celebrities from 1951 to 2004 as they appear in Google images, according to the Cross-Age Celebrity dataset (149,889 images). The data shown in panel A and B are from Rothe et al. (2018) (ref. 19) and the data in panel C is from Chen et al. (2014) (ref. 20). Dashed gold (solid blue) lines indicate the average age for female (male) faces according to each dataset. ***, p<.000001.