Pew Research Center

FOR RELEASE DECEMBER 17, 2018

Gender and Jobs in Online Image Searches

Men are overrepresented in online image searches across a majority of jobs examined; women appear lower than men in such search results for many jobs

BY *Onyi Lam, Stefan Wojcik, Brian Broderick and Adam Hughes*

FOR MEDIA OR OTHER INQUIRIES:

Stefan Wojcik, Computational Social Scientist
Kim Parker, Director, Social Trends Research
Jessica Pumphrey, Communications Associate

202.419.4372

www.pewresearch.org

RECOMMENDED CITATION

Lam, Wojcik, Broderick, and Hughes, December 2018, "Gender and Jobs in Online Image Searches"

About Pew Research Center

Pew Research Center is a nonpartisan fact tank that informs the public about the issues, attitudes and trends shaping the world. It does not take policy positions. The Center conducts public opinion polling, demographic research, content analysis and other data-driven social science research. It studies U.S. politics and policy; journalism and media; internet, science and technology; religion and public life; Hispanic trends; global attitudes and trends; and U.S. social and demographic trends. All of the Center's reports are available at www.pewresearch.org. Pew Research Center is a subsidiary of The Pew Charitable Trusts, its primary funder.

Gender and Jobs in Online Image Searches

Men are overrepresented in online image searches across a majority of jobs examined; women appear lower than men in such search results for many jobs

Online media organizations, social media sites and individuals add vast quantities of images to the web each day. These images can then appear in search engines as users look for pictures representing common phrases or topics. Because the way that men and women are represented in these online search results might be connected to the way people understand gender and society, some academic researchers have specifically focused on the ways women and men are depicted in the workplace in online images.

A new Pew Research Center study extends this line of research by using a computational method – machine vision to analyze a broad sample of images from Google Image Search that depict men and women working common jobs, and then comparing those results with real-world data about the gender composition of the U.S. workforce. The study finds that the share of each gender pictured varies widely across the spectrum of careers tested. But in the majority of jobs examined, women are somewhat underrepresented in online images relative to their actual participation rates in those jobs in the United States, based on 2017 Bureau of Labor Statistics data. Across all individuals shown in the search results, men appear 60% of the time. And, when women appear, they appear lower in the search results than men.

To conduct this analysis, Pew Research Center analyzed over 10,000 images appearing in U.S.-based, English-language search results for 105 common occupations.[1] The jobs analyzed in the study include everyday occupations such as hair stylist, librarian, butcher and plumber, among many others. Researchers used a machine vision algorithm to estimate whether each person appearing in a given image was male or female. Next, they calculated the estimated percentages of men and women depicted in the top 100 Google Image Search results for each of those jobs to assess whether the results reflected the actual percentages in each occupation who are men and women.

[1] The process of identifying the sample of jobs is described later in this report.

Key findings of the analysis include:

- **Image results for common job searches somewhat overrepresent men relative to women.** Across the sample of image searches for 105 jobs, an estimated 60% of the individuals that appeared in the image results were men (40% were women). According to Bureau of Labor Statistics (BLS) data, men made up 54% of all individuals employed in these jobs.

- **For more than half the tested job categories, images appearing in searches underrepresent women relative to their actual participation in those jobs, according to federal data.** When compared with BLS data measuring the actual share of each profession's workforce that is male and female, women were underrepresented in search results for 57% of the 105 jobs we analyzed. They were overrepresented in 42%.[2]

- **The underrepresentation of women was concentrated in a few jobs.** Across the 10 jobs where women were the most underrepresented in image searches, the average rate they appeared was 33 percentage points lower than the actual rate

The percentage of women in image search results for common jobs often differs from reality

Difference between estimated % of women in image search results and the actual % in each occupation who are women, according to BLS (selected jobs)

WOMEN **UNDERREPRESENTED** IN IMAGE SEARCH RESULTS

Job	Actual	Image search	Difference
Bill collector	71%	20%	-51
Medical records technician	92	57	-35
Bartender	57	29	-28
Probation officer	64	43	-21
General manager	34	15	-19
Chief executive	28	10	-18
Security guard	24	15	-9

WOMEN **OVERREPRESENTED** IN IMAGE SEARCH RESULTS

Job	Actual	Image search	Difference
Flight attendant	73	77	4
Physician	40	48	8
Model	78	88	10
Police	14	24	10
Computer programmer	21	34	13
Mechanic	2	24	22
Singer	38	62	24

Note: These estimates are rounded. See Methodology for more precise estimates.
Source: Pew Research Center analysis of U.S. Google Image Search data; Bureau of Labor Statistics data. Image searches were conducted July 7-Sept. 13, 2018. Images returned by searches conducted at other times may differ.
"Gender and Jobs in Online Image Searches"

PEW RESEARCH CENTER

[2] Underrepresentation and overrepresentation are defined as any estimated percentage of women in search results that is below/above the actual percentage in that occupation who are women. In image searches for "pharmacist" the rate of women returned by the search exactly matched the rate at which women held the job in reality. As a result, 1% of jobs examined here neither over- nor underrepresent women.

at which they held those jobs. Across the other 50 jobs where women were underrepresented, the average rate they appeared was just 12 points lower than their real labor force participation rate.

- **Images containing women appear further down the page in search results for many jobs.** Images depicting women engaged in jobs tended to appear lower in search results compared with men, no matter their actual share in the labor force. On average, the first image containing a woman appeared about four images from the first result, while on average the first image containing a man appeared two images from the top.

- **Image search results display *more* gender diversity than actually exists, on average.** According to the BLS, 38% of jobs in the study are predominantly held by either men or women (defined as jobs with 80% or more male or female workers). However, across image searches for all jobs, only 21% showed predominantly men or women .

What determines which images are returned in an online search?

When internet users search for images, a complex mix of factors affects what they see. Content searchers and content creators alike can influence which images draw the most attention. But researchers do not have direct access to the underlying image selection process, nor do we know how much any one factor matters when a search algorithm decides what images to select. Search algorithms are often designed to highlight images that the algorithms "expect" will be most relevant for a particular user based on their search history or other browsing behavior. But image creators and other users also can act in ways that at least partially determine the rate at which certain images appear. Algorithms are capable of learning from both content consumers and content producers. So, for example, if content producers often use men to illustrate the term "chief executive," algorithms might preserve or magnify these tendencies. At the same time, the ways that internet users click on particular images might make those images appear more relevant, and thus more likely to be selected by the algorithm. Since algorithms for image search are proprietary, the Center's research team could not evaluate them directly in this analysis.

To ensure that the images analyzed in the study depicted people at work in the jobs being searched, researchers checked whether or not the images returned in search results actually showed individuals engaged in the particular occupation of interest. Many searches for jobs did not yield a majority of images actually depicting people working in those jobs. Researchers examined each set of images and found that 44% of image search terms from a broader list of initial occupations returned a majority of images that showed people engaged in the job – rather than other people or objects associated with the job. In other words, for just 106 of the initial list of 239 jobs did the majority of images associated with each job depict the relevant worker.

The remaining 56% of searches returned images that did not reliably show a person associated with the searched occupation: Some showed clients or customers, rather than practitioners of the occupation, or depicted non-human objects. These images were not included in the analysis. For example, the majority of image results for the term "physical therapist" showed individuals receiving care rather than individuals engaging in the duties associated with being a physical therapist (see Methodology for additional details). Finally, researchers removed searches for jobs that returned fewer than 80 total images, resulting in an analysis sample of 105 sets of job-relevant images.

The image searches were conducted between July 7 and Sept. 13, 2018, and the specific images examined here may no longer appear. Images with multiple individuals were included in the analysis, and each individual was analyzed separately.

How Pew Research Center used machine vision to study gender in job searches online

Choosing which jobs to examine

Researchers started with a list of 239 jobs that had at least 100,000 employees in the U.S., acccording to the Bureau of Labor Statistics. BLS also provides information about the share of men and women employed in each job. Researchers searched Google Images for photographs of individuals engaged in these jobs and downloaded the first 100 results.

Finding relevant images

After manually inspecting all image results, researchers removed all sets of images that did not show the person engaged in the job a majority of the time. This step removed 56% of jobs that often showed individuals other than the person actually doing the work. Researchers also removed job searches that returned fewer than 80 images, or searches that had fewer than 50 faces. This resulted in a set of 105 jobs for analysis.

Gender classification

Researchers used a two-step process to identify the gender of individuals in the results. First, a machine learning algorithm detected all faces in each photograph returned in the image search. Second, researchers applied a separate algorithm to estimate the gender of particular faces. That algorithm was "trained" using a set of diverse faces and had an accuracy rate of 95% on a validation sample and 88% for a random sample of images incuded in the anlysis.

Note: Image searches were conducted July 7-Sept. 13, 2018. Images returned by searches conducted at other times may differ.
"Gender and Jobs in Online Image Searches"

PEW RESEARCH CENTER

How researchers estimated gender using machine vision

To estimate the gender of individuals in images, this analysis relies on deep learning techniques from the field of computer vision, which focuses on algorithms that identify different objects in images. Researchers used a method called "transfer learning" to repurpose an existing algorithm to the task of recognizing male and female faces.

To train a machine learning system to "learn" the facial traits associated with male and female faces, researchers assembled a set of images that humans had identified as male or female. Previous research has found that image classification algorithms available through commercial vendors suffer from inconsistent accuracy across racial groups. Accurate gender classification across racial groups is especially important in this context because researchers also used the model to classify results across different countries and languages in a separate analysis. One limitation of the machine vision model used here is that it cannot identify nonbinary individuals, since the training data included only images labeled as male or female. In addition, the model's estimate of gender is based on physical appearance, not how individuals actually identify.

The research team assembled a large and diverse set of images to train the model. Researchers used 26,981 labeled faces from different image sources to achieve a relatively balanced demographic sample from which to train the algorithm. This dataset included individuals from different countries and minority groups. The training images vary in size and image quality, which helped ensure that the classifier would work across many different kinds of images. The training images were also randomly rotated and clipped to maintain accuracy across photos in many orientations.

The gender classification model was trained with 80% of the human-labeled data, and then tested on the remaining 20% of those images. It achieved 95% accuracy on the set of images that were not used to train the model. The model's performance was broadly consistent across different minority groups (see Methodology).

After researchers tested the model's accuracy, they began applying it to search queries for particular kinds of jobs. Researchers imposed filtering criteria on the queries, in addition to checking whether the images showed individuals engaged in the job. First, researchers applied Google's "photo" filter, which is meant to remove animated or computer-generated images from search results. Second, researchers removed all the job searches with fewer than 80 images overall, or with fewer than 50 images containing human faces in the results. See Methodology for additional details.

As a final validation step, researchers tested the model on a sample of actual Google Image Search results, as classified by human coders. Across a random sample of 996 images from those included in the analysis, the model had an 88% accuracy rate. Although 1,000 images were randomly selected, four either had a dead link or were not coded by three coders.

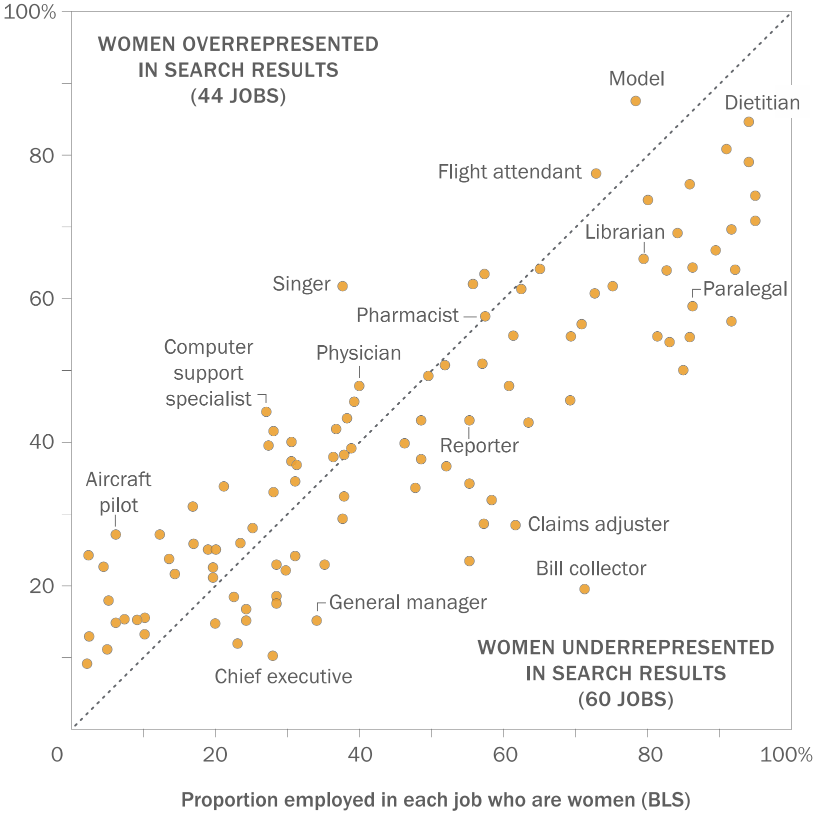
Across the 105 occupations included in the analysis, 40% of the individuals depicted across all search results were women. And for some kinds of jobs, that rate was much lower: Women appeared in search results for chief executive at a rate of 10%, and at a rate of 15% across results for general manager.

These results understate the degree to which women actually hold these jobs in the U.S. According to data from the Bureau of Labor Statistics, 28% of people employed as chief executives are women, as are 34% of those employed as general managers.[3] Across all jobs, the rate at which women appear in image searches was 6 percentage points lower than the rate at which they actually hold those jobs.

For 16% of job searches (notwithstanding *overall* rates of representation, which include any degree of over- or underrepresentation), the results were relatively close to the actual share of women holding those jobs

Women are underrepresented in image searches across 57% of jobs

Estimated % of women in image search results, by % in each occupation who are women



Note: The dotted line represents points at which women are neither under- nor overrepresented in search results.
Source: Pew Research Center analysis of U.S. Google Image Search data; Bureau of Labor Statistics data. Image searches were conducted July 7-Sept. 13, 2018. Images returned by searches conducted at other times may differ.
"Gender and Jobs in Online Image Searches"

PEW RESEARCH CENTER

[3] Among Fortune 500 companies, 4.8% of CEOs are women, according to a separate Pew Research Center study.

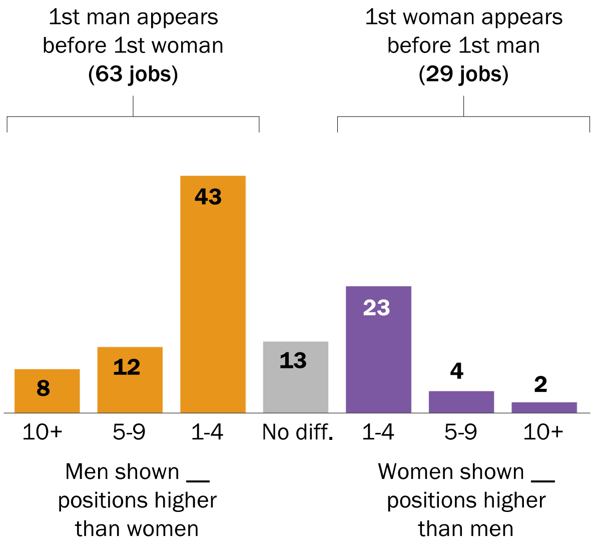
– that is, the results were within 5 percentage points of their actual share of the workforce in that occupation.

Even in jobs where women were only slightly over- or underrepresented, there are disparities in *where* images showing women appear within the search results. Researchers found that the average search *position* of an image depicting a woman tends to be lower in search results compared with images depicting a man. The higher the search position (1 being the highest, 100 the lowest), the earlier it appears in the set of images returned by Google Image Search (position is represented left to right, top to bottom). For this sample of jobs, the average image position of the first woman shown is 3.7, compared with 2.0 for the first man.

To use a concrete example, the image search for head cook (which slightly overrepresented women overall) showed the first woman in the fifth position, while a man appeared in the first position. For engineering technician (which underrepresented women by 5 percentage points), the first woman appeared in the 11th position, while the first man appeared in the first position.

Women appear lower in image search results for most jobs

Position differences between first man shown and first woman shown in Google image search results for 105 jobs



Note: For jobs with no difference, both a man and woman appeared in the same photograph.
Source: Pew Research Center analysis of U.S. Google Image search data. Image searches were conducted July 7-Sept. 13, 2018. Images returned by searches conducted at other times may differ.
"Gender and Jobs in Online Image Searches"

PEW RESEARCH CENTER

The study also found that image searches for certain gender-dominant jobs had more gender diversity relative to the actual rate at which men and women held those jobs. Researchers classified a job as *predominantly* held by men or women if at least 80% of those employed in the occupation were either men or women. Overall, 38% of these 105 jobs are predominantly held by men or women, according to BLS data. But when it came to image results for the same 105 jobs, just 21% predominantly featured men or women, using the same 80% threshold. In other words, image results for some predominantly male or female jobs are more gender diverse than would be expected if they perfectly matched the actual rate at which men and women held those jobs.

Examples of image searches that predominantly depict men include general manager and announcer, while in fact these jobs are more often held by women: 34% of general managers are women, but only 15% of individuals in search results for that occupation are women. Similarly, 23% of announcers are women, but 12% of people in search results for “announcer” are women. But for jobs like plumber, machinist and truck driver, which are predominantly held by men, image results show women at rates of 9%, 11% and 15%, respectively. These rates are substantially higher than the actual number of women who work as plumbers (2%), machinists (5%) and truck drivers (6%).

Sidebar: Gendered depictions of some jobs are consistent across countries

A separate Pew Research Center analysis approximated job-related searches for 18 different countries: Argentina, Australia, Brazil, Canada, France, Germany, India, Indonesia, Italy, Japan, Mexico, Russia, Saudi Arabia, South Africa, South Korea, Turkey, the United Kingdom and the United States. These countries were selected because they are members of the Group of 20, and together, their economies account for 63% of the global economy.[4]

Researchers obtained translations of the top 100 occupations (according to U.S. employment figures) for each of 12 languages, using either the official language of each country or the most prevalent language (if there were multiple official languages). Next, the team restricted the list to occupations that had relevant search results across a wide set of country-language combinations, using the same criteria as the U.S. analysis. For languages with gendered terms for different jobs, researchers used the male form of occupation titles when that form was the generic way to reference people of unknown gender employed in that job. For example, in Spanish, "cocinero" is a generic reference to a chef and is the masculine form of the noun. See the Methodology for additional details about the global analysis and translation process.[5]

Researchers set custom search parameters for each language and corresponding country, and then collected up to 100 images for each county-language combination.[6] Overall, researchers queried a total of 30,236 images for 307 translated job titles across the 18 countries. Some key findings of this analysis include:

- **The association between gender and certain kinds of jobs is global.** Jobs such as housekeeper, customer service representative and nurse had the highest estimated percentage of women across all languages, whereas CEO, clergy and professor had the lowest estimated percentage of women across languages, on average. Searches for nurses, for example, showed women 84% of the time in Japanese results and 80% of the time in English-language searches in the UK. Image searches for professor showed women 19% of the time for Hindi (India), compared with 22% for German (Germany).

- **Images that show women appeared lower in search results than images that show men.** The first image that showed a woman in searches for CEO or professor averaged a position of about 16 and 13, respectively, across all country-language combinations, compared with an image position

[4] The analysis excludes the European Union because some of its member states are included separately and China because Google is blocked in the country.

[5] Translations were provided by an external vendor (cApStAn). This analysis did not include all jobs across all country-language combinations, but rather, the 20 jobs for which there were at least 50 relevant (using the description defined above) results across 75% of the languages.

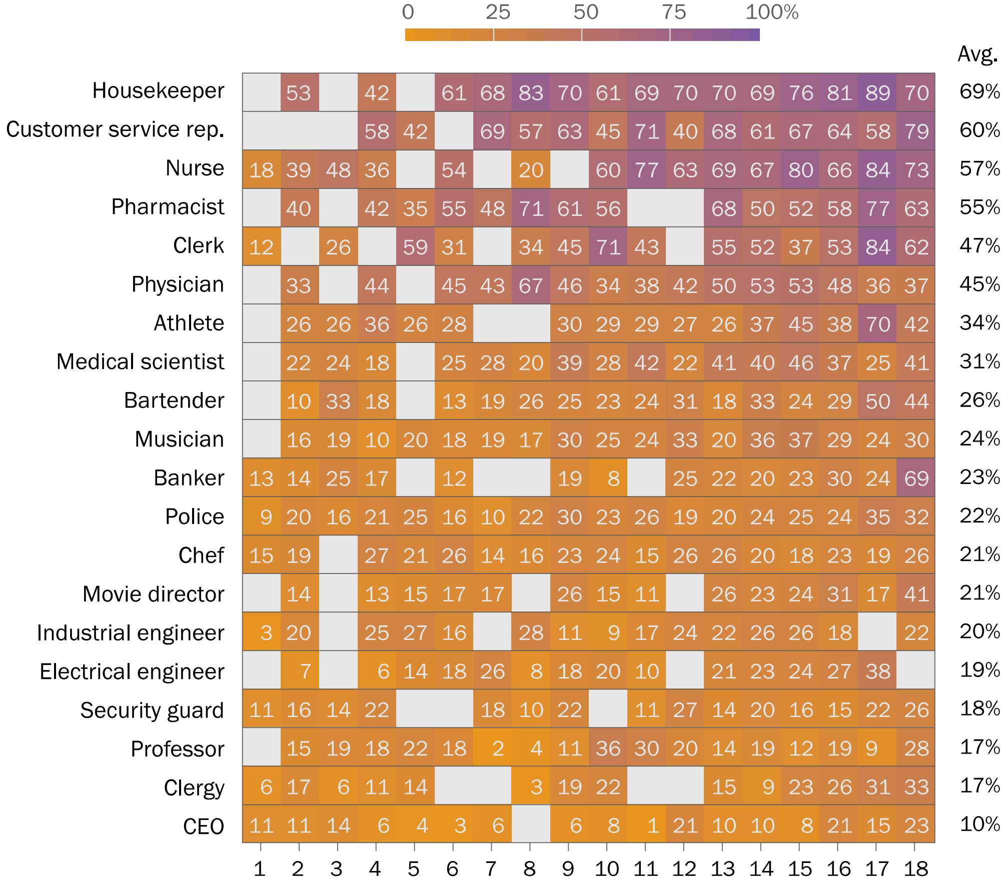
[6] Researchers used Modern Standard Arabic for the Saudi Arabia search.

of about 2 for men in both jobs. For housekeeper and customer service representative, images that depict women appear earlier in the results than those that depict men.

- **A lack of real-world comparison data for occupations at the global level limits the conclusions we can make.** Because researchers were not able to obtain parallel, real-world data about the percentages of men and women working in each job for all countries, it was not possible to conclude whether the image results overrepresent or underrepresent women in each occupation and country.

Some jobs show similar gender prevalence in image search results across languages

Estimated % of women in image search results, by job and country-language pair



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Avg.
Housekeeper		53		42		61	68	83	70	61	69	70	70	69	76	81	89	70	69%
Customer service rep.				58	42		69	57	63	45	71	40	68	61	67	64	58	79	60%
Nurse	18	39	48	36		54		20		60	77	63	69	67	80	66	84	73	57%
Pharmacist		40		42	35	55	48	71	61	56			68	50	52	58	77	63	55%
Clerk	12		26		59	31		34	45	71	43		55	52	37	53	84	62	47%
Physician		33		44		45	43	67	46	34	38	42	50	53	53	48	36	37	45%
Athlete		26	26	36	26	28			30	29	29	27	26	37	45	38	70	42	34%
Medical scientist		22	24	18		25	28	20	39	28	42	22	41	40	46	37	25	41	31%
Bartender		10	33	18		13	19	26	25	23	24	31	18	33	24	29	50	44	26%
Musician		16	19	10	20	18	19	17	30	25	24	33	20	36	37	29	24	30	24%
Banker	13	14	25	17		12			19	8		25	22	20	23	30	24	69	23%
Police	9	20	16	21	25	16	10	22	30	23	26	19	20	24	25	24	35	32	22%
Chef	15	19		27	21	26	14	16	23	24	15	26	26	20	18	23	19	26	21%
Movie director		14		13	15	17	17		26	15	11		26	23	24	31	17	41	21%
Industrial engineer	3	20		25	27	16		28	11	9	17	24	22	26	26	18		22	20%
Electrical engineer		7		6	14	18	26	8	18	20	10		21	23	24	27	38		19%
Security guard	11	16	14	22			18	10	22		11	27	14	20	16	15	22	26	18%
Professor		15	19	18	22	18	2	4	11	36	30	20	14	19	12	19	9	28	17%
Clergy	6	17	6	11	14			3	19	22			15	9	23	26	31	33	17%
CEO	11	11	14	6	4	3	6		6	8	1	21	10	10	8	21	15	23	10%

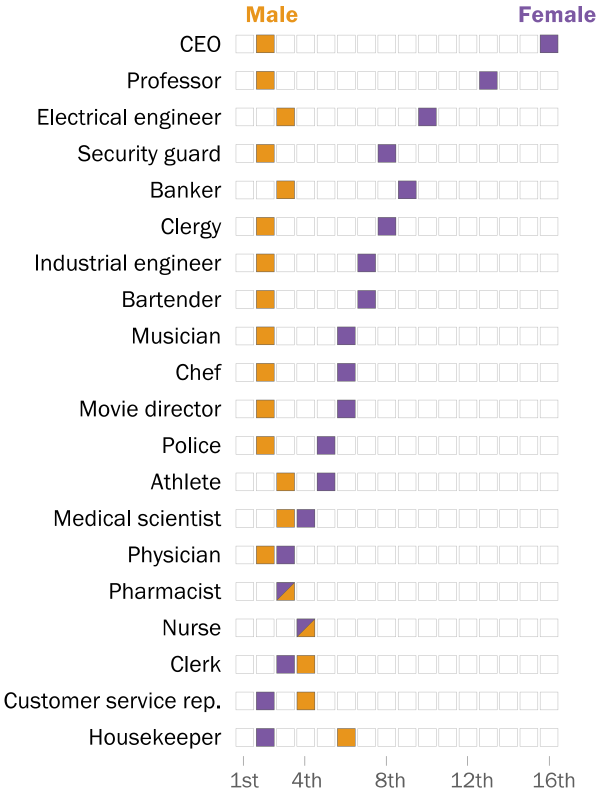
1. Arabic (Saudi Arabia)
2. Spanish (Argentina)
3. Hindi (India)
4. Spanish (Mexico)
5. German (Germany)
6. Italian (Italy)
7. Turkish (Turkey)
8. Russian (Russia)
9. English (Australia)
10. Portuguese (Brazil)
11. French (France)
12. Indonesian (Indonesia)
13. English (South Africa)
14. English (Canada)
15. English (UK)
16. English (U.S.)
17. Japanese (Japan)
18. Korean (South Korea)

Note: Gray squares indicate searches not analyzed because results either did not depict enough individuals or showed individuals not engaged in the specific occupation.
Source: Pew Research Center analysis of Google Image Search data from 18 countries. Image searches were conducted July 7-Sept. 13, 2018. Images returned by searches conducted at other times may differ.
"Gender and Jobs in Online Image Searches"

PEW RESEARCH CENTER

Women appear lower in image search results for common jobs across countries

Average position of the first man or woman in image search results for ...



Source: Pew Research Center analysis of Google Image Search data from 18 countries. Image searches were conducted July 7-Sept. 13, 2018. Images returned by searches conducted at other times may differ.
"Gender and Jobs in Online Image Searches"

PEW RESEARCH CENTER

Acknowledgments

This report is a collaborative effort based on the input and analysis of the following individuals. Find related reports online at pewresearch.org.

Primary researchers

Onyi Lam, *Computational Social Scientist*
Stefan Wojcik, *Computational Social Scientist*
Brian Broderick, *Data Science Engineer*
Adam Hughes, *Computational Social Scientist*

Research team

Emma Remy, *Data Science Assistant*
Skye Toor, *Data Science Assistant*
Patrick van Kessel, *Senior Data Scientist*
Dennis Quinn, *Data Science Analyst*

Editorial and graphic design

Peter Bell, *Design Director*
David Kent, *Copy Editor*

Communications and web publishing

Jessica Pumphrey, *Communications Associate*
Rachel Weisel, *Communications Manager*
Hannah Klein, *Communications Associate*
Travis Mitchell, *Digital Producer*

In addition, the report benefited from feedback provided by the following Pew Research Center researchers: Claudia Deane, Lee Rainie, Ruth Igielnik, Laura Silver, Kim Parker, Juliana Horowitz, Paul Hitlin, Kat Devlin and Patrick Moynihan.

Methodology

To analyze image search results for various occupations, researchers completed a four-step process. First, they created a list of U.S. occupations based on Bureau of Labor Statistics (BLS) data. Second, they translated these occupation search terms into different languages. Third, the team collected data for both the U.S. and international analysis from Google Image Search and manually verified whether or not the image results were relevant to the occupations being analyzed. Finally, researchers deployed a machine vision algorithm to detect faces within photographs, and then estimate whether those faces belong to men or women. The aggregated results of those predictions are the primary data source for this report.

Constructing the occupation list

Because researchers wanted to compare the gender breakdown in image results to real-world gender splits in occupations, the team's primary goal was to match the terms used in Google Image searches with the titles in BLS as closely as possible.

But the technical language of the BLS occupations sometimes led to questionable search results. For example, searches for "eligibility interviewers, government programs" returned images from a small number of specialized websites that actually used that specific phrase, biasing results toward those websites' images. So, the research team decided to filter out highly technical terms, using Google Trends to assess relative search popularity, relative to a reference occupation ("childcare worker").

The query selection process for the U.S. analysis involved the following steps:

1. Start with the list of BLS job titles in 2017.

2. Exclude occupations that do not have information about the fraction of women employed. For example, "credit analysts" did not have information about the fraction of women in that occupation.

3. Filter out occupations that do not have at least 100,000 workers in the U.S.

4. Remove all occupations with ambiguous job functions ("all other," "Misc.").

5. Split all titles with composite job functions into individual job titles (For example, "models and demonstrators" to "models," "demonstrators").

6. Change plural words to singular ("models" to "model") to standardize across occupations.[7]

7. Manually inspect the list to ensure that the occupations were comprehensible and likely to describe human workers. This involved removing terms that might not apply to humans (such as tester, sorter) based on the researchers' review of Google results.

8. Use Google Trends to remove unpopular or highly technical job titles. Highly technical job titles like "eligibility interviewers, government programs" are searched for less frequently than less technical titles, such as "lawyer." Accordingly, researchers decided to remove technical terms in a systematic fashion by comparing the relative search intensity of each potential job title against that of a reasonably common job title.[8] The research team compared the search intensity results for each occupation with the search intensity of "childcare worker" using U.S. search interest in 2017. Any terms with search intensity below "childcare worker" were removed from the list of job titles. The reference occupation "childcare worker" was selected after researchers manually inspected the relative search popularity of various job titles and decided that "childcare worker" was popular enough that using it as a benchmark would remove many highly technical search terms.

The global part of the analysis uses a different list of job titles meant to capture more general descriptions of the same occupations. The steps to create that list include:

1. Start with the list of BLS job titles that had at least 100,000 people working in the occupation in the U.S.

2. Remove all occupations with ambiguous job functions ("all other", "Misc.").

3. Split all titles with composite job functions into individual job titles (For example, "models and demonstrators" to "models," "demonstrators").

4. Change plural words to singular ("models" to "model") to standardize across occupations.

5. Manually inspect the list to ensure that the occupations were comprehensible and likely to describe human workers. This involved removing terms that might not apply to humans (such as tester, sorter) based on manual review of Google results.

6. Replace technical job titles with more general ones when possible to simplify translations and better represent searches. For example, instead of searching for "postsecondary

[7] For one search ("bellhops"), researchers inadvertently used the plural form of the word for the U.S. analysis.
[8] Google Trends returns results on a scale from 0 to 100, with 100 representing the highest search intensity for the terms queried within the selected region and time frame and zero the lowest.

teacher," the team searched for "professor," and instead of "chief executive," the team used "CEO."

7. Use Google Trends to filter unpopular job titles relative to a reference occupation ("childcare worker"), following the same procedure described above. Any terms with search intensity below the search intensity of the reference occupation were removed.

8. Select the top 100 terms with the most popular search intensity in Google in the U.S. within the past year.

9. Translate each job title and determine which form to use when multiple translations were available.

Translations

To conduct the international analysis, the research team chose to examine image results within a subset of G20 countries, which collectively account for 63% of the global economy. These countries include Argentina, Australia, Brazil, Canada, France, Germany, India, Indonesia, Italy, Japan, Mexico, Russia, Saudi Arabia, South Africa, South Korea, Turkey, the United Kingdom and the United States. The analysis excludes the European Union because some of its member states are included separately and China because Google is blocked in the country. Researchers used the official language of each country for each search (or the most popular language if there were multiple official languages), and worked with a translation service, cApStAn, to develop the specific search queries.

To approximate search results for each country, researchers adjusted Google's country and language settings. For example, to search jobs in India, job titles were queried in the Hindi language with the country set to India. Several countries in the study share the same official language; for example, Argentina and Mexico both have Spanish as their official language. In these cases, researchers executed separate queries for each language and country combination. The languages used in the searches were: Modern Standard Arabic, English, French, German, Hindi, Indonesian, Italian, Japanese, Korean, Portuguese, Russian, Spanish and Turkish.

Many languages spoken in these countries have gender-specific words for each occupation term. For example, in German, adding "in" to the end of the word "musiker" (musician) gives a female connotation to the word. However, the word "musiker" may not exclusively imply "male musician," and it is not the case that only male musicians can be referred to as "musiker." In consultation with the translation team, researchers identified the gender form of each job that would be used when a person of unknown gender is referenced, and searched for those terms. The male version is the default choice for most languages and occupations, but the translation team recommended using the feminine form for some cases when it was more commonly used. For example, researchers searched for “nurse” in Italian using the feminine term “infermière” rather than the masculine “infirmier” on the advice of the translation team.

In addition, some titles do not have a directly equivalent title in another language. For example, the job term “compliance officer” does not have an Italian equivalent. Finally, the same translated term can refer to different occupations in some languages. As a result, not all languages have exactly 20 search terms. Jobs lacking an equivalent translation in a given language were excluded.

Data collection

To create the master dataset used for both analyses, researchers built a data pipeline to streamline image collection, facial recognition and extraction, and facial classification tasks. To ensure that a large number of images could be processed in a timely manner, the team set up a database and analysis environment on the Amazon Web Service (AWS) cloud, which enabled the use of graphics processing units (GPUs) for faster image processing. Building this pipeline also allowed the researchers to collect additional labeled training images relatively quickly, which they leveraged to increase the diversity of the training set in advance of classifying the image search results.

Search results can be affected by the timing of the queries: Some photos could be more relevant during the time the query is executed, and therefore have a higher rank in the search results compared with searches at other times.

There are a number of filters users can apply to the images returned by Google. Under “Tools,” for example, users can signal to Google Image Search that they would like to receive images of different types, including “Face,” “Photo” and “Clip Art,” among other options. Users can also filter images by size and usage rights. For this study, researchers collected images using both the “photo” and “face” filter settings, but the results presented in this report use the “photo” filter only. Researchers made this decision because the “photo” filter appeared to provide more diverse kinds of images than the “face” filter, while also excluding clip art and animated representations of jobs.

Removing irrelevant queries

For occupations included across both the U.S. and international analysis term lists, some queries returned images that did not depict individuals engaged in the occupation being examined. Instead, they often returned images that showed clients or customers, rather than practitioners of the occupation, or depicted non-human objects. For example, the majority of image results for the term “physical therapist” showed individuals receiving care rather than individuals engaging in the duties associated with being a physical therapist.

To ensure the relevance of detected faces, researchers reviewed all of the collected images for each language, country and occupation combination. For the U.S. analysis, there were a total of 239 sets of images to review. For the international analysis, there were 1,800 sets of images to review. Queries were categorized into one of four categories based on the contents of the collected images.

- “Pass”: More than half of collected images depict only individuals employed in the queried occupation. Overall, 44% of jobs in the U.S. analysis and 43% of jobs in the global analysis fell into this category.

- “Fail”: The majority of collected images do not depict any face or depict faces irrelevant to the desired occupation. In many languages, the majority of collected images for the occupation “barber” depict only people who have been to a barber, rather than the actual barber. In the analysis of international search results, this includes queries that return images of an occupation different from that initially defined by the English translation. For example, the Arabic translation of “janitor” returns images of soccer goalies when queried in Saudi Arabia. Because the faces depicted in these images are not representative of the desired occupation, we categorize these queries as “fail.” A total of 31% of jobs in the U.S. analysis and 37% of jobs in the global analysis fell into this category.

- “Complicated”: The majority of collected images depict multiple people, some of whom are engaged in the queried occupation and some of whom are not. For example, the term “preschool teacher” and its translations often return images that feature not only a teacher but also students. These queries are categorized as “complicated” because of the difficulty in isolating the relevant faces. A total of 23% of jobs in the U.S. analysis and 17% of jobs in the global analysis fell into this category.

- “Ambiguous”: Some queries do not fall into the other categories, as there is no clear majority of image type or it is unclear whether the people depicted in the collected images are engaged in the occupation of interest. This may occur if the term has many definitions, such as “trainer,” which can refer to a person who trains athletes or various training equipment, or if the term has other usage in popular culture, such as the surname of a public figure (“baker”) or the name of a popular movie (“taxi driver”). Just 2% of jobs in the U.S. analysis and 3% of jobs in the global analysis fell into this category.

To minimize any error caused by irrelevance of detected faces in collected images, we remove all queries categorized as "fail," "complicated" or "ambiguous" and only retain those queries categorized as "pass."

Machine vision for gender classification

Researchers used a method called "transfer learning" to train a gender classifier, rather than using machine vision methods developed by an outside vendor. In some commercial and noncommercial alternative classifiers, "multitask" learning methods are used to simultaneously perform face detection, landmark localization, pose estimation, gender recognition and other face analysis tasks. The research team's classifier achieved high accuracy for the gender classification task, while allowing the research team to monitor a variety of important performance metrics.

Face detection

Researchers used the face detector from the Python library *dlib* to identify all faces in the image. The program identifies four coordinates of the face: top, right, bottom and left (in pixels). This system achieves 99.4% accuracy on the popular Labeled Faces in the Wild dataset. The research team cropped the faces from the images and stored them as separate files.

Many images collected do not contain any individuals at all. For example, all images returned by Google for the German word "Barmixer" are images of a cocktail shaker product, even with the country search parameter set to Germany. To avoid drawing inference based on a small number of images, researchers included only queries that have at least 80 images downloaded and 50 images with at least one face detected in the analysis. Across different countries, the number of faces detected in the images varied. Hindi and Indonesian had the most detected faces. This means that their images tend to feature more people in them than other languages.

The table below summarizes the number of queries, number of images and number of faces detected. Overall, researchers were able to collect over 95% of the top 100 images that we sought to download.

Training the model

Recently, research has provided evidence of algorithmic bias in image classification systems from a variety of high-profile vendors.[9] This problem is believed to stem from imbalanced training data that often overrepresents white men. For this analysis, researchers decided to train a new gender classification model using a more balanced image training set.

However, training an image classifier is a daunting task because collecting a large labeled dataset for training is very time and labor intensive, and often is too computationally intensive to actually execute. To avoid these challenges, the research team relied on a technique called "transfer learning," which involves recycling large pretrained neural networks (a popular class of machine learning models) for more specific classification tasks. The key innovation of this technique is that lower layers of the pretrained neural networks often contain features that are useful across different image classification tasks. Researchers can reuse these pretrained lower layers and fine-tune the top layers for their specific application – in this case, the gender classification task.

Image search data by country-language

Language-Country	Number of queries	Number of images	Number of faces
Arabic-Saudi Arabia	9	890	1,282
English-U.S.	20	1,961	2,457
English-UK	20	1,985	2,628
English-Australia	19	1,884	2,515
English-Canada	20	1,964	2,758
English-South Africa	20	1,952	2,721
French-France	17	1,694	1,825
German-Germany	13	1,292	1,565
Hindi-India	13	1,254	1,918
Indonesian-Indonesia	15	1,478	2,904
Italian-Italy	17	1,695	1,802
Japanese-Japan	19	1,869	2,192
Korean-South Korea	19	1,864	2,791
Portuguese-Brazil	19	1,866	1,937
Russian-Russia	16	1,567	1,653
Spanish-Argentina	18	1,777	2,201
Spanish-Mexico	19	1,855	2,101
Turkish-Turkey	14	1,389	1,719

PEW RESEARCH CENTER

The specific pretrained network researchers used is VGG16, implemented in the popular deep learning Python package *Keras*. The VGG network architecture was introduced by Karen Simonyan and Andrew Zisserman in their 2014 paper "Very Deep Convolutional Networks for Large Scale Image Recognition." The model is trained using ImageNet, which has over 1.2 million images and 1,000 object categories. Other common pretrained models include ResNet and

[9] See Buolamwini, Joy and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of Machine Learning Research.

Inception. VGG16 contains 16 weight layers that include several convolution and fully connected layers. The VGG16 network has achieved a 90% top-5 accuracy in ImageNet classification.[10]

Researchers began with the classic architecture of the VGG16 neural network as a base, then added one fully connected layer, one dropout layer and one output layer. The team conducted two rounds of training – one for the layers added for the gender classification task (the custom model), and subsequently one for the upper layers of the VGG base model.

Researchers froze the VGG base weights so that they could not be updated during the first round of training, and restricted training during this phase to the custom layers. This choice reflects the fact that weights for the new layers are randomly initialized, so if we allowed the VGG weights to be updated it would destroy the information contained within them. After 20 epochs of training on just the custom model, the team then unfroze four top layers of the VGG base and began a second round of training. For the second round of training, researchers implemented an early-stopping function. Early stopping checks the progress of the model loss (or error rate) during training, and halts training when validation loss value ceases to improve. This serves as both a timesaver and keeps the model from overfitting to the training data.

In order to prevent the model from overfitting to the training images, researchers randomly augmented each image during the training process. These random augmentations included rotations, shifting of the center of the image, zooming in/out, and shearing the image. As such, the model never saw the same image twice during training.

Selecting training images

Image classification systems, even those that draw on pretrained models, require a substantial amount of training and validation data. These systems also demand diverse training samples if they are to be accurate across demographic groups. To ensure that the model was accurate when it came to classifying the gender of people from diverse backgrounds, researchers took a variety of steps. First, the team located existing datasets used by researchers for image analysis. These include the "Labeled Faces in the Wild" (LFW) and "Bainbridge 10K U.S. Adult Faces" datasets. Second, the team downloaded images of Brazilian politicians from a site that hosts municipal-level election results. Brazil is a racially diverse country, and that is reflected in the demographic diversity in its politicians. Third, researchers created original lists of celebrities who belong to different minority groups and collected 100 images for each individual. The list of minority celebrities focused on famous black and Asian individuals. The list of famous blacks includes 22

[10] The top-5 accuracy is calculated by counting the times a predicted label matched the target label, divided by the number of data points evaluated for the five categories with the highest probabilities.

individuals: 11 men and 11 women. The list of famous Asians includes 30 individuals: 15 men and 15 women. Researchers then compiled a list of the most-populous 100 countries and downloaded up to 100 images of men and women for each nation-gender combination, respectively (for example, "French man"). This choice helped ensure that the training data included images that feature people from a diverse set of countries, balancing out the over-representativeness of white people in the training dataset. Finally, researchers supplemented this list with a set of 21 celebrity seniors (11 men and 10 women) to help improve model accuracy on older individuals. This allowed researchers to easily build up a demographically diverse dataset of faces with known gender and racial profiles.

Some images feature multiple people. To ensure that the images were directly relevant, a member of the research team reviewed each face manually and removed irrelevant or erroneous faces (e.g., men in images with women). Researchers also removed images that were too blurry, too small, and those where much of the face was obscured. In summary, the training data consist of 14,351 men and 12,630 women in images. The images belong to seven different datasets.

Training datasets

Dataset	Number of male faces	Number of female faces	Total
Bainbridge	1,023	753	1,776
Brazil Politicians	1,612	1,627	3,239
Labeled Faces in the Wild	2,839	776	3,615
Famous Blacks	755	741	1,496
Famous Asians	796	755	1,551
Country-Gender Image Search	6,629	7,335	13,964
Famous Seniors	697	643	1,340

PEW RESEARCH CENTER

Validation datasets

Dataset	Number of male faces	Number of female faces	Total
Bainbridge	221	185	406
Brazil Politicians	384	373	757
Labeled Faces in the Wild	729	214	943
Famous Blacks	176	176	352
Famous Asians	189	175	364
Country-Gender Image Search	1,685	1,799	3,484
Famous Seniors	157	141	298

PEW RESEARCH CENTER

Model performance

To evaluate whether the model was accurate, researchers applied it to a subset equivalent to 20% of the image sources: a "held out" set which was not used for training purposes. The model achieved an overall accuracy of 95% on this set of validation data. The model was also accurate on particular subsets of the data, achieving 0.96 positive predictive value on the black celebrities subset, for example.

As a final validation exercise, researched used an online labor market to create a hand-coded random sample of 996 faces. This random subset of images overrepresented men – 665 of the images were classified as male. Each face was coded by three online workers. For the 924 faces that had consensus across the three coders, the overall accuracy of this sample is 88%. Using the value 1 for "male" and 0 for "female," the precision and recall of the model were 0.93 and 0.90, respectively. This suggests that the model performs slightly worse for female faces, but that the rates of false positives and negatives was relatively low. Researchers found that many of the misclassified images were blurry, smaller in size, or obscured.

Model performance statistics

Data source	Pos. predicted value	Error rate	True positive rate	False positive rate
Bainbridge	0.978	0.022	0.962	0.018
Brazil Politicians	0.997	0.003	0.893	0.003
Labeled Faces in the Wild	0.939	0.061	0.953	0.066
Famous Blacks	0.960	0.040	0.966	0.040
Famous Asians	0.943	0.057	0.948	0.053
Country-Gender Image Search	0.899	0.101	0.869	0.029
Famous Seniors	0.964	0.036	0.957	0.032

PEW RESEARCH CENTER

Comparison of BLS data and image search results

Occupation	Bureau of Labor Statistics proportion of women	U.S. image search proportion of women	Difference
Singer	0.377	0.617	-0.240
Mechanic	0.024	0.242	-0.218
Aircraft pilot	0.062	0.271	-0.209
Maintenance worker	0.045	0.226	-0.181
Computer support specialist	0.271	0.442	-0.171
Electrical engineer	0.123	0.271	-0.148
Chemical engineer	0.169	0.310	-0.141
Judge	0.281	0.415	-0.134
Railroad conductor	0.052	0.179	-0.127
Computer programmer	0.212	0.338	-0.126
Chemical technician	0.274	0.395	-0.121
Electrician	0.025	0.129	-0.104
Police	0.136	0.237	-0.101
Paramedic	0.306	0.400	-0.094
Model	0.784	0.875	-0.091
Clergy	0.170	0.258	-0.088
Truck driver	0.062	0.148	-0.086
Construction manager	0.074	0.153	-0.079
Physician	0.400	0.478	-0.078
Radio operator	0.144	0.216	-0.072
Plumber	0.022	0.091	-0.069
Emergency medical technician	0.306	0.373	-0.067
Cook	0.393	0.456	-0.063
Retail buyer	0.558	0.620	-0.062
Machinist	0.050	0.111	-0.061
Dispatcher	0.574	0.634	-0.060
Mechanical engineer	0.092	0.152	-0.060
Distribution manager	0.192	0.250	-0.058
Shipping clerk	0.313	0.368	-0.055
Construction inspector	0.102	0.155	-0.053
Laborer	0.199	0.250	-0.051

Note: Image searches were conducted July 7-Sept. 13, 2018. Images returned by searches conducted at other times may differ.
Source: Pew Research Center analysis of Google Image Search data; Bureau of Labor Statistics data.
“Gender and Jobs in Online Image Searches”

PEW RESEARCH CENTER

Occupation	Bureau of Labor Statistics proportion of women	U.S. image search proportion of women	Difference
Chemist	0.383	0.433	-0.050
Mail clerk	0.368	0.418	-0.050
Magistrate	0.281	0.330	-0.049
Flight attendant	0.729	0.774	-0.045
Concierge	0.311	0.345	-0.034
Building inspector	0.102	0.132	-0.030
Chef	0.197	0.225	-0.028
Farmer	0.252	0.280	-0.028
Network administrator	0.235	0.259	-0.024
Athlete	0.364	0.379	-0.015
Head cook	0.197	0.211	-0.014
Stock clerk	0.379	0.382	-0.003
Computer systems analyst	0.389	0.391	-0.002
Pharmacist	0.575	0.575	0.000
Lifeguard	0.496	0.492	0.004
Customer service representative	0.651	0.641	0.010
Veterinarian	0.625	0.613	0.012
Artist	0.519	0.507	0.012
Industrial engineer	0.226	0.184	0.042
Engineering technician	0.200	0.147	0.053
Inspector	0.379	0.324	0.055
Property manager	0.486	0.430	0.056
Correctional officer	0.285	0.229	0.056
Real estate sales agent	0.571	0.509	0.062
Library assistant	0.801	0.737	0.064
Food service manager	0.463	0.398	0.065
Archivist	0.614	0.548	0.066
Bellhops*	0.311	0.241	0.070
Butcher	0.243	0.167	0.076
Film director	0.298	0.221	0.077
Musician	0.377	0.293	0.084
Security guard	0.243	0.151	0.092

Note: Image searches were conducted July 7-Sept. 13, 2018. Images returned by searches conducted at other times may differ. *Researchers inadvertently left “bellhops” plural.
Source: Pew Research Center analysis of Google Image Search data; Bureau of Labor Statistics data.
“Gender and Jobs in Online Image Searches”

PEW RESEARCH CENTER

Occupation	Bureau of Labor Statistics proportion of women	U.S. image search proportion of women	Difference
Dietitian	0.941	0.846	0.095
Restaurant hostess	0.859	0.759	0.100
Receptionist	0.910	0.808	0.102
Bailiff	0.285	0.181	0.104
Jailer	0.285	0.179	0.106
Real estate manager	0.486	0.376	0.110
Announcer	0.231	0.119	0.112
Cashier	0.727	0.607	0.120
Reporter	0.553	0.430	0.123
Janitor	0.352	0.229	0.123
Marketing specialist	0.608	0.478	0.130
Municipal clerk	0.752	0.617	0.135
Librarian	0.795	0.655	0.14
Bus driver	0.478	0.336	0.142
Physician assistant	0.709	0.564	0.145
Telemarketer	0.694	0.547	0.147
Nutritionist	0.941	0.790	0.151
Veterinary assistant	0.842	0.691	0.151
Medical scientist	0.521	0.366	0.155
Chief executive	0.280	0.102	0.178
Travel agent	0.827	0.639	0.188
General manager	0.341	0.151	0.190
Secretary	0.950	0.743	0.207
Probation officer	0.635	0.427	0.208
Correspondent	0.553	0.342	0.211
Legal assistant	0.863	0.643	0.220
Health information technician	0.917	0.696	0.221
Licensed practical nurse	0.895	0.667	0.228
Clinical laboratory technician	0.693	0.458	0.235
Administrative assistant	0.950	0.708	0.242
Tax collector	0.584	0.319	0.265
File clerk	0.814	0.547	0.267

Note: Image searches were conducted July 7-Sept. 13, 2018. Images returned by searches conducted at other times may differ.
Source: Pew Research Center analysis of Google Image Search data; Bureau of Labor Statistics data.
"Gender and Jobs in Online Image Searches"

PEW RESEARCH CENTER

Occupation	Bureau of Labor Statistics proportion of women	U.S. image search proportion of women	Difference
Paralegal	0.863	0.589	0.274
Nurse practitioner	0.922	0.640	0.282
Bartender	0.573	0.286	0.287
Office clerk	0.831	0.539	0.292
Restaurant host	0.859	0.546	0.313
News analyst	0.553	0.234	0.319
Claims adjuster	0.617	0.284	0.333
Medical records technician	0.917	0.568	0.349
Interviewer	0.850	0.500	0.350
Bill collector	0.713	0.195	0.518

Note: Image searches were conducted July 7-Sept. 13, 2018. Images returned by searches conducted at other times may differ.
Source: Pew Research Center analysis of Google Image Search data; Bureau of Labor Statistics data.
"Gender and Jobs in Online Image Searches"

PEW RESEARCH CENTER