HAL open science

What Is Beautiful Continues to Be Good

Maria Matsangidou, Jahna Otterbacher

To cite this version:

Maria Matsangidou, Jahna Otterbacher. What Is Beautiful Continues to Be Good. 17th IFIP Conference on Human-Computer Interaction (INTERACT), Sep 2019, Paphos, Cyprus. pp.243-264, 10.1007/978-3-030-29390-1_14 . hal-02877678

HAL Id: hal-02877678

https://inria.hal.science/hal-02877678v1

Submitted on 22 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

What is Beautiful Continues to be Good:

People Images and Algorithmic Inferences on Physical Attractiveness

Maria Matsangidou[1][0000-0003-3804-5565] and Jahna Otterbacher[1,2][0000-0002-7655-7118]

[1]Research Centre on Interactive Media, Smart Systems and Emerging Technologies (RISE), Nicosia, Cyprus, {m.matsangidou, j.otterbacher}@rise.org.cy

[2] Cyprus Center for Algorithmic Transparency, Open University of Cyprus, Nicosia, Cyprus

Abstract. Image recognition algorithms that automatically tag or moderate content are crucial in many applications but are increasingly opaque. Given transparency concerns, we focus on understanding how algorithms tag people images and their inferences on *attractiveness*. Theoretically, attractiveness has an evolutionary basis, guiding mating behaviors, although it also drives social behaviors. We test image-tagging APIs as to whether they encode biases surrounding attractiveness. We use the Chicago Face Database, containing images of diverse individuals, along with subjective norming data and objective facial measurements. The algorithms encode biases surrounding attractiveness, perpetuating the stereotype that "what is beautiful is good." Furthermore, women are often misinterpreted as men. We discuss the algorithms' reductionist nature, and their potential to infringe on users' autonomy and well-being, as well as the ethical and legal considerations for developers. Future services should monitor algorithms' behaviors given their prevalence in the information ecosystem and influence on media.

Keywords. Algorithmic bias, Attractiveness, Image recognition, Stereotyping

1 Introduction

Image recognition is one of the success stories of modern machine learning and has become an indispensable tool in our information ecosystem. Beyond its early applications in more restricted domains such as military (e.g., satellite imagery), security and surveillance, or medical imaging, image recognition is an increasingly common component in consumer information services. We have become accustomed to seamlessly organizing and/or searching collections of images in real time - even those that lack descriptive metadata. Similarly, the technology has become essential in fields such as interactive marketing and campaigning, where professionals must learn about and engage target audiences, who increasingly communicate and share in a visual manner.

Underpinning the widespread "democratization" of image recognition is that it has enjoyed rapid technical progress in recent years. Krizhevsky and colleagues [34] first

introduced the use of deep learning for image classification, in the context of the ImageNet Challenge.[1] This work represented a significant breakthrough, improving classification accuracy by over 50% as compared to previous state-of-the-art algorithms. Since then, continual progress has been made, e.g., in terms of simplifying the approach [49] and managing the required computational resources [55].

Nevertheless, along with the rapid uptake of image recognition, there have been high-profile incidents highlighting its potential to produce socially offensive – even discriminatory – results. For instance, a software engineer and Google Photos user discovered in 2015 that the app had labeled his photos with the tag "gorillas." Google apologized for the racial blunder, and engineers vowed to find a solution. However, more than three years later, only "awkward workarounds," such as removing offensive tags from the database, have been introduced.[2] Similarly, in late 2017, it was widely reported that Apple had offered refunds to Asian users of its new iPhone X, as its Face ID technology could not reliably distinguish Asian faces. The incident drew criticism in the press, with some questioning whether the technology could be considered racist.[3]

Given the rise of a visual culture that dominates interpersonal and professional exchanges in networked information systems, it is crucial that we achieve a better understanding of the biases of image recognition systems. The number of cognitive services (provided through APIs) for image processing and understanding has grown dramatically in recent years. Without a doubt, this development has fueled creativity and innovation, by providing developers with tools to enhance the capabilities of their software. However, as illustrated in the above cases, the social biases in the underlying algorithms carry over to applications developed.

Currently, we consider whether tagging algorithms make inferences on *physical attractiveness.* There are findings suggesting that online and social media are projections of the idealized self [13] and that the practice of uploading pictures of one's self represents a significant component of self-worthiness [52]. At the same time, some researchers have argued that the media culture's focus on physical appearance and users' repeated exposure to this is correlated to increased body image disturbance [41]. This was further supported by a recent study [5] suggesting that individuals with attractive profile photos in dating websites, are viewed more favorably and with positive qualities.

The role of physical appearance in human interaction cannot be denied; it is the first personal characteristic that is observable to others in social interactions. These obvious characteristics are perceived by others and in response, shape ideas and beliefs, usually based on cultural stereotypes. Fig. 1 presents three images from the Chicago Face Database (CFD) as well as the respective demographic information, subjective ratings and physical measures provided in the CFD [38]. In the last row, the output of four tagging algorithms is shown. As illustrated, while tagging algorithms are not designed to be discriminatory, they often output descriptive tags that reflect prevalent conceptions and even stereotypes concerning attractiveness. For instance, the woman on the left, who

[1] http://www.image-net.org/challenges/LSVRC/

[2] https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/

[3]https://www.newsweek.com/iphone-x-racist-apple-refunds-device-cant-tell-chinese-people-apart-woman-751263

CFD judges found to be the most attractive and feminine of the three persons (see "Attractive" score), is described with tags such as "cute" and "smile," while the others are described more neutrally. Furthermore, the less attractive woman is described as a man. The question is to what extent this happens systematically.

Black Woman, 29.8 years Attractive: 5.08 Babyface: 1.76 Feminine: 5.63 Masculine: 1.52 Nose length: 250 Nose width: 268 Lip thickness: 162	Latino Man, 35.5 years Attractive: 1.54 Babyface: 2.21 Feminine: 1.17 Masculine: 4.71 Nose length: 182 Nose width: 230 Lip thickness: 104	White Woman, 24.9 years Attractive: 3.11 Babyface: 1.99 Feminine: 3.86 Masculine:3.01 Nose length: 251 Nose width: 188 Lip thickness: 65
Clarifai: woman, Afro, dreadlock, cute Microsoft: hairpiece, clothing, wear, smile Watson: person, woman, female Imagga: afro, attractive, pretty, model	Clarifai: man, casual, cool, friendly Microsoft: person, necktie, wearing, shirt Watson: stubble, coonskin cap, afro hair style Imagga: man, face, male, person, creation	Clarifai: face, man, casual, eye Microsoft: man, looking, shirt, wearing Watson: person, pompadour hairstyle, skin Imagga: person, face, man, male, handsome

Fig. 1. Example people images with tags from four APIs and Chicago Face Database subjective ratings/physical measures.

We approach the study of "algorithmic attractiveness" through the theoretical lens of *evolutionary social psychology.* This reductionist approach attempts to explain why certain physical characteristics are correlated to human judgments of attractiveness, from an evolutionary perspective (i.e., explaining sexual selection) [12]. As early as 1972, Dion and colleagues [19] discovered a commonly held belief: "what is beautiful is good". This led researchers to describe a stereotype, that attractive people are often perceived as having positive qualities. Many researchers have validated this finding and have suggested that the stereotype is applied across various social contexts [21].

While these theories are not without controversy, we find the evolutionary perceptive particularly interesting in light of algorithmic biases. While people use dynamic and non-physical characteristics when judging attractiveness, algorithms may make subjective judgments based on one static image. These reductionist judgments, an "algorithmic attractiveness," carry over into applications in our information ecosystem (e.g., image search engines, dating apps, social media) as illustrated in Fig. 2. From media

effects theories (e.g., cultivation theory [26], social learning theory [6]) we know that with repeated exposure to media images (and judgments on those images) we come to accept these depictions as reality. Thus, there is a real danger of exacerbating the bias; whoever is deemed to be "beautiful and good" will continue to be perceived as such.

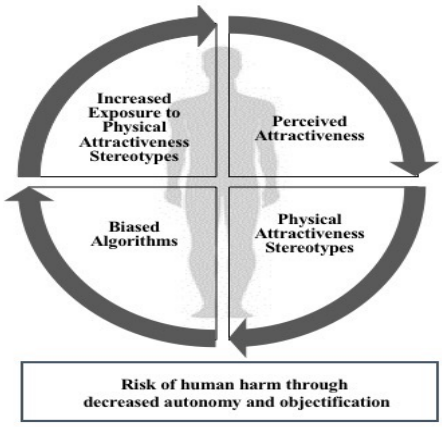

Fig. 2. Conceptual model of bias perpetuation.

In light of the need for greater algorithmic transparency and accountability [18], and the growing influence of image recognition technology in the information landscape, we test popular tagging algorithms to answer three research questions:

RQ1) Do the algorithms exhibit evidence of evolutionary biases with respect to human physical attractiveness?

RQ2) Do the taggers perpetuate the stereotype that attractive people are "good," having positive social qualities?

RQ3) Do the algorithms infer gender accurately when tagging images of men and women of different races?

2 Related Work

We provide a review of evolutionary social psychology theory that provides an explanation for the human biases surrounding physical attractiveness. In addition, we explain the potential for social consequences – the so-called "physical attractiveness stereotype." Finally, having seen that gender and physical attractiveness are closely related, we briefly review recent work on gender-based biases in image recognition algorithms.

2.1 Evolutionary roots of physical attractiveness

Interpersonal relationships are essential for healthy development. However, subjective differences occur when it comes to choosing/rejecting another person. *Physical appearance* plays a key role in the development of interpersonal attraction, the positive attitudes or evaluations a person holds for another [3, 30]. Walster and colleges provided

an influential example in the early 60s [58]. In their study, they matched individuals based on personality, intelligence, social skills, and physical attractiveness. Results revealed that only attractiveness mattered in terms of partners liking one another.

Evolutionary theory has long been used to explain the above. As Darwin [16] claimed, sexual selection of the preferable mate is influenced by aesthetic beauty. In humans, research suggests that heterosexual men tend to select a partner based on physical attractiveness, more so than women. This is because of reproductive value, the extent to which individuals of a given age and sex will contribute to the ancestry of future generations [22]. As a result, men show a particular interest in women of high fertility. In women, reproductive value typically peaks in the mid-teens and declines with age [56]. Therefore, women's physical attractiveness is usually associated with youthful characteristics, such as smooth skin, good muscle tone, lustrous hair, a small waste-to-hip ratio, and babylike features such as big eyes, a small nose and chin, and full lips [10, 11, 14, 50, 53, 54, 60].

Physical appearance is less important for heterosexual women. Women are believed to be concerned with the external resources a man can provide, such as his earning capacity, ambition, and industriousness [10]. This is because male parental investment tends to be less than female parental investment. Specifically, a copulation that requires minimal male investment can produce a 9-month investment by the woman that is substantial in terms of time, energy, resources, and foreclosed alternatives. Therefore, women tend to select men who can provide them with the resources necessary to raise their child [57, 60]. These characteristics might be inferred from a man's physical appearance and behaviors; favorable traits include being tall, displaying dominant behavior, having an athletic body, big cheekbones, a long and wide chin, and wearing fashionable clothing that suggests high social status [11].

RQ1 asks whether there is evidence of evolutionary biases in the tags output by algorithms to describe input person images. As will be explained, the dataset we use provides human (subjective) ratings on a large set of people images, as well as physical facial measurements; therefore, we will compare human and algorithmic attractiveness judgments on the depicted persons, in relation to the above-described findings.

2.2 Physical attractiveness stereotype

The physical attractiveness stereotype ("what is beautiful is good") is the tendency to assume that people who are attractive also possess socially desirable traits [19, 42]. A convincing example was given by a study where a photo of either a "beautiful" or an "unattractive" woman was given to a heterosexual man, along with biographical information [51]. Men tended to be positively biased toward a "beautiful" woman, who was perceived as being a sociable, poised, humorous, and skilled person. This was in stark contrast to the perceptions expressed by men who were given an "unattractive" woman's picture. "Unattractive" women were expected to be unsociable, awkward, serious, and socially inept. In a follow-up analysis, a blind telephone call between the man and the depicted woman was facilitated. The expectations of the men had a dramatic impact on their behavior. Specifically, the men who spoke with an "attractive

woman" exhibited a response and tone of voice evaluated as warm and accepting. However, those who spoke with an "unattractive" woman used a less inviting tone [51].

Many studies have highlighted the tendency to believe that "what is beautiful is good" [19, 35, 42]. Furthermore, research has shown that several stereotypes link personality types to a physically attractive or unattractive appearance. Specifically, individuals in casual acquaintances invariably assume that an attractive person is more sincere, noble and honest, which usually results in pro-social behavior towards the attractive person in comparison to a non-attractive person [19, 21]. However, these stereotypes are not consistent worldwide. Cultural differences may affect the perception of beauty stereotype. Research has shown that the content of an attractive stereotype depends on cultural values [59] and is changing over time [31].

To answer **RQ2**, we will examine the extent to which humans and tagging algorithms tend to associate depicted persons deemed to be physically attractive with more positive traits/words, as compared to images of less attractive individuals.

2.3 Gender and image recognition

Most research to date on social biases in image recognition has come from outside HCI and has focused on technical solutions for specific gender biases. For instance, Zhao and colleagues [62] documented gender-based activity biases (e.g., associating verbs such as cooking/driving with women/men) in image labels in the MS-COCO dataset [37]. They introduced techniques for constraining the corpus that mitigate the bias resulting in models trained on the data.

Looking more specifically at face recognition in images, Klare and colleagues [33] warned that certain demographic groups (in particular, people of color, women, and young people ages 18 to 30) were systematically more prone to recognition errors. They offered solutions for biometric systems used primarily in intelligence and law enforcement contexts. Introna and Wood's results [32] confirms these observations. More recent work [36] applied deep learning (convolutional neural networks (CNNs)) to age and gender recognition. They noted that while CNNs have brought about remarkable improvements in other image recognition tasks, "accurately estimating the age and gender in unconstrained settings…remains unsolved."

In addition, the recent work by Buolamwini and Gebru [8] found not only that gender classification accuracy in popular image recognition algorithms is correlated to skin color, but also that common training data sets for image recognition algorithms tend to over-/under-represent people with light/dark skin tones. Furthermore, computer vision researchers are taking seriously the issue of diversity in people image processing, as evidenced by new initiatives such as the InclusiveFaceNet [46].

To address **RQ3**, we consider the extent to which image recognition algorithms produce tags that imply gender and examine the accuracy of the implied gender.

3 Methodology

3.1 Image recognition algorithms

We study four-image recognition APIs – Clarifai, Microsoft Cognitive Services' Computer Vision, IBM Watson Visual Recognition and the Imagga Auto-Tagger. All are easy-to-use; we made minimal modifications to their Python code examples to upload and process images. While three of the companies offer specific models for face recognition, we experiment with the general tagging models. First, tagging algorithms are more broadly used in information services and applications (e.g., image sharing sites, e-commerce sites) to organize and/or moderate content across domains. Face recognition algorithms infer specific, but more limited, attributes (e.g., age, gender) whereas more general tools often make additional inferences about the depicted individual and who he or she is (e.g., doctor, model), often using subjective tags (e.g., attractive, fine-looking). Finally, face recognition algorithms are less mature and, as mentioned previously, not yet very accurate. Next, we provide descriptions of each tagger. As proprietary tools, none provides a specific explanation of how the tags for an input image are chosen (i.e., they are opaque to the user/developer). As these services are updated over time, it is important to note that the data was collected in July – August 2018.

Clarifai[4] describes its technology as a "proprietary, state-of-the-art neural network architecture." For a given image, it returns up to 20 descriptive concept tags, along with probabilities. The company does not provide access to the full set of tags; we used the general model, which "recognizes over 11,000 different concepts." Microsoft's Computer Vision API[5] analyzes images in various ways, including content-based tagging as well as categorization. We use the tagging function, which selects from over 2,000 tags to provide a description for an image. The Watson Visual Recognition API[6] "uses deep learning algorithms" to analyze an input image. Its default model returns the most relevant classes "from thousands of general tags." Finally, Imagga's tagger[7] combines "state-of-the-art machine learning approaches" with "semantic processing" to "refine the quality of the suggested tags[8]." Like Clarifai, the tag set is undefined; however, Imagga returns all associated tags with a confidence score (i.e., probability). Following Imagga's suggested best practices, we retain all tags with a score greater than 0.30.

[4] https://clarifai.com/technology
[5] https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/home#tagging-images
[6] https://www.ibm.com/watson/developercloud/visual-recognition/api/v3/curl.html?curl#introduction
[7] https://docs.imagga.com
[8] https://imagga.com/technology/auto-tagging.html

3.2 Person images: the Chicago Face Database (CFD)

The CFD [38] is a free resource[9] consisting of 597 high-resolution, standardized images of diverse individuals between the ages of 18 and 40 years (see Table 1). It is designed to facilitate research on a range of psychological phenomena, (e.g., stereotyping and prejudice, interpersonal attraction). Therefore, it provides extensive data about the depicted individuals (see Fig. 1 for examples). The database includes both subjective norming data and objective physical measurements (e.g., nose length/width), on the pictures.[10] At least 30 independent judges, balanced by race and gender, evaluated each CFD image.[11] The questions for the subjective norming data were posed as follows: "Now, consider the person pictured above and rate him/her with respect to other people of the same race and gender. For example, if the person was Asian and male, consider this person on the following traits relative to other Asian males in the United States. - **Attractive** (1-7 Likert, 1 = Not at all; 7 = Extremely)". Fifteen additional traits were evaluated, including Babyface, Dominant, Trustworthy, Feminine, and Masculine.

For our purposes, a significant benefit of using the CFD is that the individuals are depicted in a similar, neutral manner; if we were to evaluate images of people collected "in the wild", we would have images from a variety of contexts with varying qualities. In other words, using the CFD enables us to study the behavior of the tagging algorithms in a controlled manner.

Table 1: Mean/median attractiveness of depicted individuals.

	Asian	Black	Latino/a	White
Women	N=57; 3.64/3.62	N=104; 3.33/3.15	N=56; 3.81/3.56	N=90; 3.45/3.39
Men	N=52; 2.85/2.85	N=93; 3.17/3.12	N=52; 2.94/2.90	N=93; 2.96/2.96

3.3 Producing and analyzing image tags

The CFD images were uploaded as input to each API. Table 2 summarizes the total number of tags output by each algorithm, the unique tags used and the most frequently used tags by each algorithm. For all taggers other than Imagga, we do not observe subjective tags concerning physical appearance amongst the most common tags. However, it is interesting to note the frequent use of tags that are interpretive in nature; for instance, Watson often labels people as being "actors" whereas both Microsoft and Watson frequently interpret the age group of depicted persons (e.g., "young," "adult"). Imagga's behavior clearly deviates from the other taggers, in that three of its most frequent tags refer to appearance/attractiveness.

[9] http://faculty.chicagobooth.edu/bernd.wittenbrink/cfd/index.html

[10] Physical measurements are reported in pixels (i.e., are measured from photos.)

[11] See [38] for additional details. Scores by individual judges are not provided; the CFD contains the mean scores over all judges.

Table 2: Tags generated by the APIs across 597 images.

	Clarifai	Microsoft	Watson	Imagga
Total	11,940	12,137	3,668	6,772
Unique	95	74	72	54
10 most frequent	Portrait, one, people, isolated, casual, looking, look, eye, man, face	White, shirt, wearing, standing, posing, young, black, smiling, glasses, looking	Person, light brown color, people, ash grey, coal black, face, stubble, adult person, actor, woman	Portrait, face, person, handsome, man, male, beard, adult, attractive, model

We post-processed the output tags using the Linguistic Inquiry and Wordcount (LIWC) tool [45]. LIWC is a collection of lexicons representing psychologically meaningful concepts. Its output is the percentage of input words that map onto a given lexicon. We used four concepts: female/male references, and positive/negative emotion.

Table 3 provides summary statistics for the use of tags related to these concepts. We observe that all four taggers use words that reflect gender (e.g., man/boy/grandfather vs. woman/girl/grandmother). While all taggers produce subjective words as tags, only Clarifai uses tags with negative emotional valence (e.g., isolated, pensive, sad).

Finally, we created a custom LIWC lexicon with words associated with physical attractiveness. Three native English speakers were presented with the combined list of unique tags used by the four algorithms (a total of 220 tags). They worked independently and were instructed to indicate which of the words could be used to describe a person's physical appearance in a positive manner. This yielded a list of 15 tags: *attractive, casual, cute, elegant, fashionable, fine-looking, glamour, masculinity, model, pretty, sexy, smile, smiling, trendy, handsome.* There was full agreement on 13 of the 15 tags; two of three judges suggested "casual" and "masculinity." As shown in

Table 3, the Watson tagger did not use any tags indicating physical attractiveness. In addition, on average, the Microsoft tagger output fewer "attractiveness" tags as compared to Clarifai and Imagga.

Table 3: Mean/median percentages (%) of tags reflecting LIWC concepts.

	Clarifai	Microsoft	Watson	Imagga
Total tags	20.00/20	20.33/20	11.31/11	11.69/12
Female ref.	2.49/0	1.56/0	5.02/0	.01 /0
Male ref.	4.66/4.75	4.26/4.76	2.47/0	17.73/21.43
Positive	14.0/14.3	4.58/4.76	.31/0	11.14/9.09
Negative	7.52/9.5	N/A	N/A	N/A
Attractiveness	14.9/15	4.58/4.76	N/A	18.49/16.67

3.4 Detecting biases in tag usage

To summarize our approach, the output descriptive tags were interpreted by LIWC through its lexicons. LIWC scores were used in order to understand the extent to which

each algorithmic tagger used gender-related words when describing a given image, or whether word-tags conveying positive or negative sentiment were used. Likewise, our custom dictionary allowed us to determine when tags referred to a depicted individual's physical attractiveness. As will be described in Section 4, we used appropriate statistical analyses to then compare the taggers' descriptions of a given image to those assigned manually by CFD judges (i.e., subjective norming data described in Section 3.2). In addition, we evaluated the taggers' outputs for systematic differences as a function of the depicted individuals' race and gender, to explore the tendency for tagging bias.

4 Results

We examined the perceived attractiveness of the individuals depicted in the CFD, by humans and tagging algorithms. Based on theory, perceived attractiveness as well as the stereotypes surrounding attractiveness, differ considerably by gender. Therefore, analyses are conducted separately for the images of men and women, and/or control for gender. Finally, we considered whether the algorithms make correct inferences on the depicted person's gender.

Table 4 summarizes the variables examined in the analysis. In some cases, for a CFD variable, there is no corresponding equivalent in the algorithmic output. Other times, such as in the case of gender, there is an equivalent, but it is measured differently. For clarity, in the tables below, we indicate in the top row the origin of the variables being analyzed (i.e., CFD or algorithmic output).

Table 4: Summary of variables examined and (type).

	CFD	Algorithmic output
Gender	Self-reported (nominal)	Use of gendered tags (continuous)
Age	Self-reported (continuous)	N/A
Race	Self-reported (nominal)	N/A
Physical characteristics	Facial measurements (continuous)	N/A
Attractiveness	Human-rated (ordinal)	Use of attractiveness tags (continuous)
Other subjective (e.g., Trustworthy)	Human-rated (ordinal)	Use of positive/negative emotion tags (continuous)

4.1 Evolutionary biases and physical attractiveness

4.1.1 Human judgments

We examined whether the CFD scores on physical attractiveness are consist with the evolutionary social psychology findings. Gender is highly correlated to physical attractiveness. Over all images, judges associate attractiveness with femininity (r = .899, $p<.0001$). However, for images of men, attractiveness is strongly related to masculinity (r = .324, $p<.0001$), which is negatively correlated to women's attractiveness (r = -.682, $p<.0001$). Similarly, men's attractiveness is positively correlated to being per-

ceived as "Dominant" ($r = .159$, $p<.0001$), where the reverse is true of women's attractiveness ($r = -.219$, $p<.0001$). For both genders, "Babyface" features and perceived youthfulness (-age) are correlated to attractiveness in the same manner. The Pearson correlation analyses are presented in Table 5.

Table 5: Correlation between perceived attractiveness and youthful characteristics by gender in the CFD[12].

	CFD Gender					
CFD judgments	Men			Women		
	1	2	3	1	2	3
1. Attractive	-	-.24***	.18***	-	-.32***	.22***
2. Age		-	-.68***		-	-.69***
3. Babyface			-			-

4.1.2 Algorithmic judgments

Next, we assessed if algorithms encode evolutionary biases. In parallel to the observation that CFD judges generally associate femininity with attractiveness, Table 6 examines correlations between algorithms' use of gendered tags and attractiveness tags. Clarifai and Imagga behavior is in line with CFD judgments. For both algorithms, there is a negative association between the use of masculine and attractiveness tags, while feminine and attractiveness tags are positively associated in Clarifai output.

Table 6: Correlations between algorithms' use of "attractiveness" tags and gendered tags.

	Gendered Tags	
	Female	Male
Clarifai	.32***	-.28***
Microsoft	.05	-.04
Watson	N/A	N/A
Imagga	.06	-.45***

While Table 6 examined correlations of two characteristics of algorithmic output, Table 7 examines correlations between the algorithms' use of attractiveness tags and three CFD characteristics. We observe a significant correlation between Clarifai and Imagga's use of "attractiveness" tags, and the human judgments on attractiveness. These two algorithms exhibit the evolutionary biases discussed, with a positive correlation between indicators of youthfulness and attractiveness. The Microsoft tagger shows a reverse trend. However, it may be the case that its tags do not cover a broad enough range of words to capture the human notion of attractiveness; there is a positive, but insignificant, correlation between the CFD Attractiveness and attractiveness tags.

Since Clarifai and Imagga exhibited the most interesting behavior, a more in-depth analysis was carried out on them to see which static, physical characteristics correlate

[12] We use the following conventions for statistical significance in tables: *$p<.05$, **$p<.01$, ***$p<.001$.

to the use of attractiveness tags. A separate Pearson correlation was conducted for images of men and women, in terms of their physical facial measurements in the CFD and the two algorithms' use of "attractiveness" tags (Table 8). The analysis revealed a strong positive correlation for men between attractiveness and having a wide chin. Both genders revealed a positive correlation between attractiveness and luminance; once again, this feature can be considered a signal of youth.

Table 7: Correlations between algorithms' use of "attractiveness" tags and CFD Attractiveness.

	CFD judgments		
Algorithmic attractiveness tags	Attractiveness	Babyface	Age
Clarifai	.25***	.06	-.18***
Microsoft	.05	-.12**	.11**
Watson	N/A	N/A	N/A
Imagga	.32***	.11***	-.17***

Finally, a strong positive correlation was observed between attractiveness and nose length for women. This result, along with the negative correlation between women's attractiveness and nose width and shape highlights the relationship between the existence of a small nose and the perception of attractiveness. We should also point out that for both genders, attractive appearance is correlated with petite facial characteristics, such as light lips, small eyes.

Table 8: Correlations between facial measurements and algorithms' use of attractiveness tags.

		Static, physical facial measurements from CFD											
		1	2	3	4	5	6	7	8	9	10	11	12
Men	1.Attractive (Clarifai)	-	.49***	-.24***	.10	-.23***	-.24***	-.26***	.03	-.01	.01	-.23***	.17***
	Attractive (Imagga)		.38***	-.32***	.12*	-.30***	-.17**	-.14**	.01	-.04	.03	.07	.12*
	2.Luminance		-	-.58***	.32**	-.60***	-.58***	-.60***	-.06	-.04	-.06	-.47***	.37***
	3.Nose Width			-	-.16**	.75***	.58***	.54***	-.05	-.14*	-.14*	.16***	-.28***
	4.Nose Length				-	-.76***	-.19***	-.29***	.32***	.23***	.19***	-.32***	-.12*
	5.Nose shape					-	.51***	.55***	-.24***	-.24***	-.21***	.31***	-.10
	6.Lip Thickness						-	.97***	.15*	.07	.05	.19***	-.51***
	7.Lip Fullness							-	.10	.05	.09	.23***	-.50***
	8.Eye Height								-	.89***	.93***	.00	-.18***
	9.Eye Width									-	.07	-.23**	.00
	10.Eye Shape										-	-.03	-.21***
	11.Eye Size											-	-.13*
	12.Chin Length												-

Women	1. Attractive (Clarifai)	-	.40***	-.24***	.20***	-.31***	-.16***	-.22***	.02	-.18***	-.06	-.05	.11
	Attractive (Imagga)		.32***	-.06	.18**	-.17**	-.05	-.07	-.01	-.01	.01	-.02	-.01
	2.Luminance		-	-.52***	.31***	-.57***	-.49***	-.57***	.01	-.31***	.02	-.07	.35***
	3.Nose Width			-	-.04	.71***	.60***	.49***	-.05	.04	-.17**	-.20***	-.32***
	4.Nose Length				-	-.72***	-.05	-.24***	.42***	-.43***	.29***	.17***	-.21***
	5.Nose shape					-	.43***	.50***	-.32***	.32***	-.32***	-.26***	-.05
	6.Lip Thickness						-	.94***	.04	-.01	-.06	-.08	-.54***
	7.Lip Fullness							-	-.02	.12*	-.04	.02	-.52***
	8.Eye Height								-	-.05	.87***	.89***	-.23***
	9.Eye Width									-	.04	.31***	-.07
	10.Eye Shape										-	.87***	-.23***
	11.Eye Size											-	-.19***
	12.Chin Length												-

In conclusion, we observe that both humans (i.e., CFD judges) and algorithms (particularly Clarifai and Imagga), associate femininity, as well as particular static facial features with attractiveness. Furthermore, there is a strong correlation between CFD indicators of attractiveness and the use of attractiveness tags by algorithms.

4.1 Social stereotyping and physical attractiveness

We examined whether the attractiveness stereotype is reflected in the CFD judgments, and next, whether this is true for the algorithms as well. Table 9 details the correlations between perceived attractiveness and the other subjective attributes rated by CFD judges. The first six attributes refer to perceived emotional states of the persons whereas as the last two are perceived character traits. There is a clear correlation between the perception of attractiveness and positive emotions/traits, for men and women.

Table 9: Correlation between Attractiveness and other subjective attributes in the CFD.

	CFD Gender	
CFD subjective norming	Men	Women
Afraid	-.99	-.25***
Angry	-.27***	-.35***
Disgusted	-.26***	-.37***
Happy	.45***	.48***
Sad	-.28***	-.43***
Surprised	.04	.03
Threatening	-.23***	-.40***
Trustworthy	.55***	.59***

We do not always have equivalent variables in the CFD and the algorithmic output. Therefore, to examine whether algorithms perpetuate the stereotype that "what is beautiful is good," we considered the use of LIWC positive emotion words in tags, as a function of the known (CFD) characteristics of the images. Images were divided into two groups (MA/more attractive, LA/less attractive), separated at the median CFD score (3.1 out of 7). Table 10 presents an ANOVA for each algorithm, in which attractiveness (MA vs. LA), gender (M/W) and race (W, B, L, A) are used to explain variance in the use of positive tags. For significant effects, effect size (η^2) is in parentheses.

Table 10: ANOVA to explain variance in the use of positive emotion tags.

	CFD characteristics (demographic information and attractiveness rating)							
	Attractive (A)	Gender (G)	Race (R)	A*G	A*R	G*R	A*G*R	Sig. Diff.
Clarifai	6.08* (.0006)	24.26*** (.04)	4.93** (.02)	4.17* (.006)	.38	3.91** (.02)	1.85	A: MA>LA G: W>M R: A,W>B G*R: AW>BW,BM,WM WW>AM,BW,LM, WM
Microsoft	5.05* (.006)	.065	3.16* (.02)	.02	1.38	1.18	.72	A: MA>LA R: W>L
Watson	6.327* (.004)	9.71** (.02)	6.54*** (.03)	1.15	2.28	.77	.70	A: MA > LA G: W>M R: W>L L>B
Imagga	5.526* (.003)	31.76*** (.01)	16.00*** (.03)	6.45* (.001)	.88	4.19** (.004)	.18	A: MA > LA G: W>M R: A>W B>L,W

The right-most column reports significant differences according to the Tukey Honestly Significant Differences test. All three taggers tend to use more positive words when describing more versus less attractive individuals. However, while the main effect on "attractiveness" is statistically significant, its size is very small. It appears to be the case that the depicted person's gender and race play a greater role in explaining the use of positive tags. Clarifai and Watson describe women more positively than men. Clarifai describes images of Blacks less positively than images of Asians and Whites, while Watson favors Latinos/as over Blacks, but Whites over Latinos/as. The Microsoft tagger, which shows no significant main effect on gender, favors Whites over Latinos/as.

4.2 Gender (mis)inference

Although the tagging algorithms studied do not directly perform gender recognition, they do output tags that imply gender. We considered the use of male/female reference words per the LIWC scores. Table 11 shows the percent of images for which a gender

is implied (% Gendered). In particular, we assume that when an algorithm uses tags of only one gender, and not the other, that the depicted person's gender is implied. This is the case in 80% of the images processed by Clarifai and Imagga, whereas Microsoft and Watson-produced tags imply gender in almost half of the images. Implied gender was compared against the gender recorded in the CFD; precision appears in parentheses. As previously mentioned, the Imagga algorithm used female reference tags only in the case of one woman; its strategy appears to be to use only male reference words.

Table 11: Use of gendered tags (per LIWC) and precision on implied genders.

	Algorithmically implied gender				
	% Gendered	Male only	Female only	Male & Female	No gendered tags
Clarifai	80%	358 (.81)	122 (1.00)	117	0
Microsoft	42%	245 (.53)	5 (1.00)	178	169
Watson	44%	141 (.86)	124 (.97)	0	329
Imagga	78%	465 (.61)	1 (1.00)	0	131

The algorithms rarely tag images of men with feminine references (i.e., there is high precision when implying that someone is a woman). Only three images of men were implied to be women, and only by Watson. In contrast, images of women were often tagged incorrectly (i.e., lower precision for inferring that someone is a man). Table 12 breaks down the gender inference accuracy by the depicted person's race. Cases of "no gendered tags" are considered errors. The results of the Chi-Square Test of Independence suggest that there is a relationship between race and correct inference of gender, for all algorithms other than Imagga. For these three algorithms, there is lower accuracy on implied gender for images of Blacks, as compared to Asians, Latino/as, and White.

Table 12: Proportion of correctly implied genders, by person's self-reported race in the CFD.

	Race reported in the CFD					
	Overall accuracy	Asian (n=109)	Black (n=197)	Latino/a (n=108)	White (n=183)	χ^2
Clarifai	.69	.75	.59	.76	.72	14.35**
Microsoft	.23	.28	.10	.26	.31	29.34***
Watson	.41	.44	.31	.49	.45	12.96**
Imagga	.47	.47	.47	.47	.48	.09

5. Discussion and Implications

5.1 Discussion

Consumer information services, including social media, have grown to rely on computer vision applications, such as taggers that automatically infer the content of an input image. However, the technology is increasingly opaque. Burrell [9] describes three types of opacity, and the taggers we study exhibit all three. First, the algorithms are

proprietary, and their owners provide little explanation as to how they make inferences about images or even the set of all possible tags (e.g., Clarifai, Imagga). In addition, because all are based on deep learning, it may be technically infeasible to provide meaningful explanations and, even if they were provided, few people are positioned to understand an explanation (technical illiteracy). In short, algorithmic processes like image taggers have become "power brokers" [18]; they are delegated many everyday tasks (e.g., analyzing images on social media, to facilitate our sharing and retrieval) and operate largely autonomously, without the need for human intervention or oversight [61]. Furthermore, there is a tendency for people to perceive them as objective [27, 44] or even to be unaware of their presence or use in the system [20].

As our current results demonstrate, image tagging algorithms are certainly *not objective* when processing images depicting people. While we found no evidence that they output tags conveying negative judgments on physical appearance, positive tags such as "attractive," "sexy," and "fine-looking" were used by three algorithms in our study; only Watson did not output such descriptive tags. This is already of concern, as developers (i.e., those who incorporate APIs into their work) and end users (i.e., those whose images get processed by the APIs in the context of a system they use) might not expect an algorithm designed for tagging image content, to produce such subjective tags. Even more telling is that persons with certain physical characteristics were more likely to be labeled as attractive than others; in particular, the Clarifai and Imagga algorithms' use of such tags was strongly correlated to human judgements of attractiveness.

Furthermore, all four algorithms associated images of more attractive people (as rated by humans), with tags conveying positive emotional sentiment, as compared to less attractive people, thus reinforcing the physical attractive stereotype, "beautiful is good." Even when the depicted person's race and gender were controlled, physical attractiveness was still related to the use of more positive tags. The significant effects on the race and gender of the depicted person, in terms of explaining the use of positive emotion tags, is also of concern. Specifically, Clarifai, Microsft and Watson tended to label images of whites with more positive tags in comparison to other racial groups, while Clarifai, Watson and Imagga favored images of women over men.

As explained, the theoretical underpinnings of our study are drawn from evolutionary social psychology. These theories are *reductionist* in nature – they rather coldly attempt to explain interpersonal attraction as being a function of our reproductive impulses. Thus, it is worrying to observe correlations between "algorithmic attractiveness" and what would be predicted by theory. In a similar vein, Hamidi and colleagues [29] described automatic gender recognition (AGR) algorithms, which extract specific features from the input media in order to infer a person's gender, as being more about "gender reductionism" rather than recognition. They interviewed transgender individuals and found that the impact of being misgendered by a machine was often even more painful than being misrecognized by a fellow human. One reason cited was the perception that if algorithms misrecognizes them, this would solidify existing standards in society.

As depicted in Fig. 2, we fear that this reductionism could also influence standards of physical attractiveness and related stereotypes. As mentioned, while offline, people use other dynamic, non-physical cues in attractiveness judgments, image tagging algorithms do not. Algorithms objectify static people images with tags such as "sexy" or

"fine-looking." Given the widespread use of these algorithms in our information ecosystem, algorithmic attractiveness is likely to influence applications such as image search or dating applications, resulting in increased circulation of people images with certain physical characteristics over others, and could lead to an increased stereotypical idea of the idealized self. Research on online and social media has shown that the online content that is uploaded by the users can enhance and produce a stereotypical idea of perfectionism [13] that in many cases leads to narcissistic personality traits [15, 28, 39].

In addition, there is some evidence suggesting that media stereotypes have a central role in creating and exacerbating body dissatisfaction. Constant exposure to "attractive" pictures in media enhances comparisons between the self and the depicted ideal attractive prototype, which in return creates dissatisfaction and 'shame' [25, 41, 48]. In some cases, the exposure has significant negative results for mental health such as the development of eating disorders since the user shapes projections of the media idealized self [2, 7, 43]. One can conclude that the continuous online exposure to attractive images that are tagged by algorithms with positive attributes may increase these stereotypical ideas of idealization with serious threats to users' mental health.

With specific reference to gender mis-inference, it is worth noting again that certain applications of AGR are not yet mature; as detailed in the review of related work, AGR from images is an extremely difficult task. Although general image tagging algorithms are not meant to perform AGR, our study demonstrated that many output tags do imply a depicted person's gender. On our dataset, algorithms were much more likely to misimply that women were men, but not vice versa. The application of these algorithms, whether specifically designed for AGR or a general image-tagging tool, in digital spaces, might negatively impact users' sense of *human autonomy* [24].

Being labelled mistakenly as a man, like the right-most woman in Fig. 1, or not being tagged as "attractive," when images of one's friends have been associated with such words, could be a painful experience for users, many of whom carefully craft a desired self-presentation [4]. Indeed, the prevalence of algorithms in social spaces has complicated self-presentation [17], and research has shown that users desire a greater ability to manage how algorithms profile them [1]. In short, our results suggest that the use of image tagging algorithms in social spaces, where users share people images, can pose a danger for users who may already suffer from having a negative self-image.

5.2 Implications for developers

Third party developers increasingly rely on image tagging APIs to facilitate capabilities such as image search and retrieval or recommendations for tags. However, the opaque nature of these tools presents a concrete challenge; any inherent biases in the tagging algorithms will be carried downstream into the interactive media developed on these. Beyond the ethical dimensions of cases such as those described in Section 1, there are also emerging legal concerns related to algorithmic bias and discrimination. In particular, the EU's new General Data Protection Regulation - GDPR, will affect the routine use of machine learning algorithms in a number of ways.

For instance, Article 4 of the GDPR defines profiling as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain

aspects relating to a natural person[13]." Developers will need to be increasingly sensitive to the potential of their media to inadvertently treat certain groups of users unfairly and will need to implement appropriate measures as to prevent discriminatory effects. In summary, developers – both those who provide and/or consume "cognitive services" such as image tagging algorithms - will need to be increasingly mindful of the quality of their output. Because of the lack of transparency in cognitive services may have multiple sources (economic, technical) future work should consider the design of independent services that will monitor them for unintended biases, enabling developers to make an informed choice as to which tools they use.

5.3 Limitations of the study

We used a theoretical lens that was reductionist in nature. This was intentional, as to highlight the reductionist nature of the algorithms. However, it should be noted that the dataset we used, the CFD, also has some limiting characteristics. People depicted in the images are labeled strictly according to binary gender, and their respective races are reported as a discrete characteristic (i.e., there are no biracial people images). Gender was also treated as a binary construct in the APIs we examined. Nonetheless, the current study offered us the opportunity to compare the behavior of the tagging algorithms in a more controlled setting, which would not be possible if we had used images collected in the wild. In future work, we intend to expand the study in terms of both the datasets evaluated, and the algorithms tested. It is also true that algorithms are updated from time-to-time, by the owners of the cognitive services; therefore, it should be noted the temporal nature and the constant updating of the machine learning driving the API's propose another limitation for the study. In future work we shall improve the study with the newest versions of the APIs to process the images.

6. Conclusion

This work has contributed to the ongoing conversation in HCI surrounding AI technologies, which are flawed from a social justice perspective, but are also becoming intimately interwoven into our complex information ecosystem. In their work on race and chatbots, Schlesinger and colleagues [47] emphasized that "neat solutions" are not necessarily expected. Our findings on algorithmic attractiveness and image tagging brings us to a similar conclusion. We have highlighted another dimension upon which algorithms might lead to discrimination and harm and have demonstrated that image tagging algorithms should not be considered objective when it comes to their interpretation of people images. But what can be done? Researchers are calling for a paradigm shift; *Diversity Computing* [23] could lead to the development of algorithms that mimic the ways in which humans learn and change their perspectives. Until such techniques are feasible, HCI researchers and practitioners must continue to scrutinize the opaque tools that tend to reflect our own biases and irrationalities.

[13] https://gdpr-info.eu/art-4-gdpr/

Acknowledgements

This research has been funded by the European Union's Horizon 2020 research and innovation programme under grant agreements No. 739578 (RISE), 810105 (CyCAT) and the Government of the Republic of Cyprus (RISE).

References

1. Alvarado, O, & Waern, A.: Towards Algorithmic Experience: Initial Efforts for Social Media Contexts. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, pp. 286-298 (2018).
2. Agras, W. S., & Kirkley, B. G.: Bulimia: Theories of etiology. Handbook of eating disorders, 367-378, (1986).
3. Aron, A., Lewandowski, G. W., Mashek, Jr., D., & Aron, E. N.: The self-expansion model of motivation and cognition in close relationships." The Oxford Handbook of Close Relationships, 90-115, (2013).
4. Birnholtz, J., Fitzpatrick, C., Handel, M., & Brubaker, J. R.: Identity, identification and identifiability: the language of self-presentation on a location-based mobile dating app. Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services (MobileHCI '14), pp. 3-12, (2014).
5. Brand, R. J., Bonatsos, A., D'Orazio, R., & DeShong, H.: What is beautiful is good, even online: Correlations between photo attractiveness and text attractiveness in men's online dating profiles. Computers in Human Behavior, 28(1), 166-170, (2012).
6. Brown, J. D.: Mass media influences on sexuality. Journal of Sex Research, 39(1), 42-45, (2002).
7. Brownell, K. D.: Personal responsibility and control over our bodies: When expectation exceeds reality. Health Psychology, 10(5), 303-310, (1991)
8. Buolamwini, J., & Gebru, T.: Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on Fairness, Accountability and Transparency, pp. 77-91, (2018).
9. Burrell, J.: How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms. Big Data & Society, 1-12, (2016).
10. Buss, D. M.: Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. Behavioral and brain sciences, 12(1), 1-14, (1989).
11. Buss, D. M.: The Evolution of Desire. Basic Books, (1994).
12. Buss D. M., & Kenrick, D. T.: Evolutionary social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), The Handbook of Social Psychology, 982-1026. New York, NY, US: McGraw-Hill (1998).
13. Chua, T. H. H., & Chang, L.: Follow me and like my beautiful selfies: Singapore teenage girls' engagement in self-presentation and peer comparison on social media. Computers in Human Behavior, 55, 190-197 (2016).
14. Cunningham, M. R.: Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. Journal of personality and social psychology, 50(5), 925 -935, (1986).
15. Davenport, S. W., Bergman, S. M., Bergman, J. Z., & Fearrington, M. E.: Twitter versus Facebook: Exploring the role of narcissism in the motives and usage of different social media platforms. Computers in Human Behavior, 32, 212-220, (2014).

16. Darwin, C.: The Descent of Man, and Selection in Relation to Sex. London, UK: John Murray, (1871).
17. DeVito, M. A., Birnholtz, J., Hancock, J. T., French, M., & Liu, S.: How People Form Folk Theories of Social Media Feeds and What it Means for How We Study Self-Presentation. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), Paper 120.
18. Diakopoulos, N.: Accountability in algorithmic decision making. Commun. ACM 59, 2, 56-62, (2016).
19. Dion, K., Berscheid E., & Walster, E.: What is beautiful is good. Journal of personality and social psychology, 24(3), 285-290, (1972).
20. Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K. Hamilton, K., & Sandvig, C.: "I always assumed that I wasn't really that close to [her]": Reasoning about Invisible Algorithms in News Feeds. In Proceedings of the 33rd Annua17.
21. Feingold, A.: Good-looking people are not what we think. Psychological Bulletin 111, 2, 304-341, (1992).
22. Fisher, R. A.: The Genetical Theory of Natural Selection. Clarendon Press, (1930).
23. Fletcher-Watson, S., De Jaegher, H., van Dijk, J., Frauenberger, C., Magnée, M., & Ye, J.: Diversity computing. Interactions 25, 5, 28-33, (2018).
24. Friedman, B.: Value-sensitive design. Interactions 3,6, 16-23, (1996).
25. Garner, D. M., & Garfinkel, P. E.: Socio-cultural factors in the development of anorexia nervosa. Psychological medicine, 10(4), 647-656, (1980).
26. Gerbner, G., Gross, L., Morgan, M., Signorielli, N., & Shanahan, J.: Growing up with television: Cultivation processes. Media effects: Advances in theory and research, 2, 43-67, (2002).
27. Gillespie, T.: The Relevance of Algorithms. In Media Technologies: Essays on communication, materiality and society. Tarleton Gillespie, Pablo Boczkowski and Kirsten Foot (eds.) MIT Press, Cambridge, MA, 167-193, (2014).
28. große Deters, F., Mehl, M. R., & Eid, M.: Narcissistic power poster? On the relationship between narcissism and status updating activity on Facebook. Journal of Research in Personality, 53, 165-174, (2014).
29. Hamidi, F., Scheuerman, M. K., & Branham, S. M.: Gender Recognition or Gender Reductionism? The Social Implications of Embedded Gender Recognition Systems. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM,
30. Hatfield E., & Rapson, R. L.: Love, sex, and intimacy: Their psychology, biology, and history. HarperCollins College Publishers, (1993).
31. Hatfield, E., & Sprecher, S.: Mirror, Mirror: The Importance of Looks in Everyday Life. SUNY Press, (1986)
32. Introna, L., & Wood, D.: Picturing algorithmic surveillance: The politics of facial recognition systems. Surveillance & Society 2, 177-198, (2004)
33. Klare, B. F., Burge, M. J., Klontz, J. C., Bruegge, R. W. V., & Jain, A. K.: Face recognition performance: Role of demographic information. IEEE Transactions on Information Forensics and Security, 7(6), 1789-1801, (2012).
34. Krizhevsky, A. Sutskever, I., & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (NIPS), 1097-1105, (2012).
35. Langlois, J. H.: From the eye of the beholder to behavioral reality: Development of social behaviors and social relations as a function of physical attractiveness. In Physical appearance, stigma, and social behavior: The Ontario Symposium, 3, 23-51, Lawrence Erlbaum, (1986).

36. Levi, G., & Hassner, T.: Age and gender classification using convolutional neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, (2015).
37. Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dolla´r, P., & Zitnick, C. L.: Microsoft coco: Common objects in context. In European Conference on Computer Vision, pages 740–755. Springer, (2014)
38. Ma, D. S., Correll J., & Wittenbrink, B.: The Chicago face database: A free stimulus set of faces and norming data. Behavior research methods, 47(4), 1122-1135, (2015).
39. McKinney, B. C., Kelly, L., & Duran, R. L.: Narcissism or openness?: College students' use of Facebook and Twitter. Communication Research Reports, 29(2), 108-118, (2012).
40. McLellan B., & McKelvie S. J.: Effects of age and gender on perceived facial attractiveness. Canadian Journal of Behavioural Science, 25(1), 135-143, (1993).
41. Meier, E. P., & Gray, J.: Facebook photo activity associated with body image disturbance in adolescent girls. Cyberpsychology, Behavior, and Social Networking, 17(4), 199-206, (2014).
42. Miller, A. G.: Role of physical attractiveness in impression formation. Psychonomic Science, 19(4), 241-243, (1970).
43. Morris, A., Cooper, T., & Cooper, P. J.: The changing shape of female fashion models. International Journal of Eating Disorders, 8(5), 593-596, (1989).
44. O'Neil, C.: Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Random House, (2016).
45. Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K.: The development and psychometric properties of LIWC2015. https://utexasir.tdl.org/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf'sequence=3, (2015).
46. Ryu, H. J., Adam, H., & Mitchell, M.: Inclusivefacenet: Improving face attribute detection with race and gender diversity. ICML'18 FATML Workshop, Stockholm, Sweden (2017)
47. Schlesinger, A., O'Hara, K. P., Taylor, A. S.: Let's Talk About Race: Identity, Chatbots, and AI. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, (2018).
48. Silberstein, LR, Striegel-Moore, RH, & Rodin, J.: Feeling fat: A woman's shame. In HB Lewis (Ed.), The role of shame in symptom formation (pp. 89-108). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc, (1987).
49. Simonyan, K., & Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, (2014).
50. Singh, D.: Adaptive significance of female physical attractiveness: role of waist-to-hip ratio. Journal of personality and social psychology, 65(2), 293-307, (1993).
51. Snyder, M., Tanke E. D., & Berscheid, E.: Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. Journal of Personality and social Psychology, 35(9), 656-672, (1977).
52. Stefanone, M. A., Lackaff, D., & Rosen, D.: Contingencies of self-worth and social-networking-site behavior. Cyberpsychology, Behavior, and Social Networking, 14(1-2), 41-49, (2011).
53. Stroebe, W., Insko, C. A., Thompson, V. D., & Layton, B. D.: Effects of physical attractiveness, attitude similarity, and sex on various aspects o interpersonal attraction. Journal of Personality and Social Psychology, 18(1), 79-91, (1971).
54. Symons, D.: The Evolution of Human Sexuality. Oxford University Press, (1979).

55. Szegedy, C., Wei, L., Yangqing, J., Pierre, S., Scott, R., Dragomir, A., Dumitru, E., Vincent, V., & Andrew, R.: Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, 1-9, (2015).
56. Thornhill R., & Alcock, J.: The Evolution of Insect Mating Systems. Harvard University Press, (1983).
57. Trivers, R. L.: Parental investment and sexual selection. In: Sexual selection and the descent of man, ed. B. Campbell. Aldine, (1972).
58. Walster, E., Aronson, V., Abrahams, D., & Rottman, L.: Importance of physical attractiveness in dating behavior. Journal of Personality and Social Psychology, 4(5), 508-516, (1966).
59. Wheeler, L., Kim, Y.: What is beautiful is culturally good: The physical attractiveness stereotype has different content in collectivistic cultures. Personality and Social Psychology Bulletin, 23(8), 795-800, (1997).
60. Williams, G. C.: Pleiotropy, natural selection and the evolution of senescence. Evolution 11:398-411, (1957).
61. Wilson, M.: Algorithms (and the) everyday. Information Communication & Society 20,1: 137-150, (2017).
62. Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K. W.: Men also like shopping: Reducing gender bias amplification using corpus-level constraints. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 2979– 2989 Copenhagen, Denmark, September, 20, 7–11, (2017)