Gender stereotypes in AI-generated images

Francisco-José García-Ull; Mónica Melero-Lázaro

Francisco-José García-Ull ✉
https://orcid.org/0000-0002-7999-4807

Universidad Europea de Valencia
Facultad de Ciencias Sociales
Pg. de l'Albereda, 7
46010 València, Spain
franciscojose.garcia@universidadeuropea.es

Mónica Melero-Lázaro
https://orcid.org/0000-0002-2517-8468
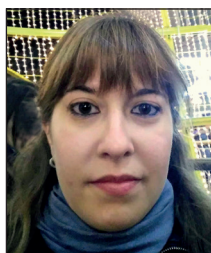
Universidad de Valladolid
Facultad de Filosofía y Letras
Pza. del Campus Universitario
47011 Valladolid, Spain
monimeler@outlook.com

Abstract

This study explores workplace gender bias in images generated by *DALL-E 2*, an application for synthesising images based on artificial intelligence (AI). To do this, we used a stratified probability sampling method, dividing the sample into segments on the basis of 37 different professions or prompts, replicating the study by Farago, Eggum-Wilkens and Zhang (2020) on gender stereotypes in the workplace. The study involves two coders who manually input different professions into the image generator. *DALL-E 2* generated 9 images for each query, and a sample of 666 images was collected, with a confidence level of 99% and a margin of error of 5%. Each image was subsequently evaluated using a 3-point Likert scale: 1, not stereotypical; 2, moderately stereotypical; and 3, strongly stereotypical. Our study found that the images generated replicate gender stereotypes in the workplace. The findings presented indicate that 21.6% of AI-generated images depicting professionals exhibit full stereotypes of women, while 37.8% depict full stereotypes of men. While previous studies conducted with humans found that gender stereotypes in the workplace exist, our research shows that AI not only replicates this stereotyping, but reinforces and increases it. Consequently, while human research on gender bias indicates strong stereotyping in 35% of instances, AI exhibits strong stereotyping in 59.4% of cases. The results of this study emphasise the need for a diverse and inclusive AI development community to serve as the basis for a fairer and less biased AI.

Keywords

Artificial intelligence; AI; *OpenAI*; *DALL-E*; Synthetic images; AI-generated images; Imaging; Gender stereotypes; Sex biases; Gender biases; Gender differences; Professions; Workers; Jobs; Ethics; Discrimination; Inequalities; Fairness; Equity.

1. Introduction

Artificial Intelligence (AI) has significantly influenced society in recent years, transforming task execution and decision-making across diverse fields, such as legal (**Sourdin**, 2018), health (**Loftus** *et al*., 2020), business (**Nica** *et al*., 2022; **Belhadi** *et al*., 2022) and education (**Porayska-Pomsta**; **Rajendran**, 2019). However, concerns have arisen regarding potential gender biases in AI systems.

One issue contributing to gender bias in AI is the lack of diversity in the data used to train the models. If the data used to train an AI model lacks sufficient representation of the population's diversity, the model may inadvertently learn and

perpetuate gender stereotypes (**Leavy**, 2018). One of the recent developments in AI is the *DALL-E 2* (2021) language generator, developed by *OpenAI*. Although *DALL-E 2* has demonstrated impressive content generation capabilities, acknowledging and addressing potential manifestations of gender biases in its use is essential. *https://openai.com/dall-e-2*

The model's training data primarily originates from the internet, resulting in the potential inclusion of gender stereotypes prevalent in online culture. Taking into account these types of limitations when designing a rigorous study is important.

So is accounting for how gender biases might manifest in the data generated by *DALL-E 2*. For instance, a study by **Buolamwini** and **Gebru** (2018) found that commercial facial recognition systems had a higher error rate in classifying individuals of colour and gender non-binary individuals, indicating that AI systems can perpetuate existing gender biases in society. Due to the swift adoption of applications like *DALL-E 2*, *ChatGPT*, and other *OpenAI* systems (**Vincent**, 2020), recognising and addressing the gaps and limitations that perpetuate models and stereotypes becomes essential. In recent years, various publications, including articles, books, and documentaries, have highlighted the presence of gender, race, and class biases in new digital technologies (**D'Ignazio**; **Klein**, 2020; **Criado-Pérez**, 2020).

To mitigate gender biases in *DALL-E 2*, establishing stereotype-free databases to train the model has been suggested. This would reduce biases and foster gender equality. Further suggestions include using debiasing techniques to eliminate gender biases in the results generated by *DALL-E 2* (**Caliskan**; **Bryson**; **Narayanan**, 2017).

Considering possible gender biases in its use is important. AI systems have the potential to replicate prevailing gender stereotypes in society if the data used to train the models lacks sufficient diversity (**Leavy** *et al.*, 2020). Addressing these biases requires the implementation of data diversification and debiasing techniques. Continued research and monitoring of potential gender biases in *DALL-E 2* and other AI systems, along with their associated benefits and risks (**De-Carvalho**, 2021), are vital. These systems will have diverse applications and may directly or indirectly influence our interpersonal relationships. Ensuring fairness, equity, and ethical usage of these systems is paramount (**Cortina**, 2019).

2. Theoretical framework

2.1. *OpenAI* and generative Artificial Intelligence

Artificial Intelligence has made significant strides in recent decades, thanks to the combination of machine learning techniques and the enhanced capacity for data processing. One of the most recent developments in this field is *DALL-E*, an AI generator developed by *OpenAI* that is capable of generating images based on text descriptions (**Brown** *et al.*, 2020). The recently released mini version of *DALL-E*, known as *DALL-E 2*, has demonstrated its impressive capability to generate highly accurate and detailed images from basic text descriptions.

DALL-E 2 is an upgraded version of its predecessor, *DALL-E*, released in 2021. In contrast to its predecessor, *DALL-E 2* has a more streamlined architecture and reduced data processing capacity, making it more accessible and user-friendly for a wide range of users and applications. Despite this smaller size, *DALL-E 2* has proven to be just as effective at generating images from text descriptions.

In terms of how it works, *DALL-E 2* employs generative deep learning and Generative Adversarial Network (GAN) technology, where a neural network is trained to generate images based on input data. In the case of *DALL-E 2*, the input data comprises a text description, which is then processed by the neural network to produce a corresponding image. This technique relies on machine learning principles and the neural network's capacity to discern patterns and associations within the input data.

Regarding *DALL-E 2*'s potential applications, one of the most obvious is its use in advertising and graphic design. *DALL-E 2* could be used to generate tailor-made images for advertising campaigns or to craft distinctive image designs for various products. Each newly generated image from this system is original and unpublished, which raises notable controversy and presents author copyright limitations (**Estupiñán-Ricardo** *et al.*, 2021). It can also be applied to education, by enabling the generation of images to illustrate concepts in textbooks or class presentations. It is in this area where certain studies have highlighted gender biases in the depiction of women in science-related contexts (**Manassero**; **Vázquez**, 2003; **Francescutti**, 2018). The judicious application of AI has the potential to transcend these limitations, contributing to a more equitable and inclusive society. *DALL-E 2* is also applied in the animation industry and video game production, as it facilitates the automated generation of scenarios and characters.

Nonetheless, there are ethical concerns regarding this technology (**Quirós-Fons**; **García-Ull**, 2022). One of the main apprehensions is the potential use of *DALL-E 2* to produce false or misleading content. There are also privacy and security concerns, as *DALL-E 2* could be used to generate images of individuals without their consent or used as a tool for digital violence (**Pérez-Gómez** *et al.*, 2020). Concurring with this, **Veliz** (2021) emphasises the influential power that the manipulation of private data can yield, underscoring the need to foster initiatives and tools that safeguard user privacy. Also stressing the centralisation of power motivated by technological hegemony, authors such as **Crawford** (2021) assert the emergence of a trend towards

> One issue contributing to gender bias in AI is the lack of diversity in the data used to train the models

greater inequality. They call upon technology companies to harness AI to steer the trajectory towards democratic values and the restructuring of the political and social landscape. In a similar vein, **O'Neil** (2018) cautions against the presence of opaque and unregulated algorithms and models that perpetuate discrimination, favouring the fortunate while penalising the marginalised.

2.2. GAN technology

Generative Adversarial Networks (GAN) are machine learning models capable of generating new and realistic content such as images, audio, and text. These models are made up of two neural networks: the first is called a generator and the second a discriminator. The generator is responsible for generating new content, while the discriminator is responsible for determining whether the generated content is real or false. The two models engage in a zero-sum game, competing against each other with the objective of enhancing the quality of the generated content.

The most influential work on GANs was carried out by **Goodfellow** *et al.* (2014) in their article "*Generative Adversarial Networks*". In this paper, the authors presented a basic GAN architecture and showed how it could be used to generate images of human faces. Since then, GANs have been used to generate a variety of content, including images, audio, and text.

One of the primary advantages of GANs is their ability to generate realistic content, which has resulted in their application across various domains, such as video game production, animation, and product design. GANs have also been used to generate scenarios and characters for video games, and product images that are used in decision making for design.

GANs have been applied in the medical field too. They have been used to generate brain CT images to aid the diagnosis of neurodegenerative diseases (**Laino** *et al*., 2022), and of cells and tissues to assist in the research and development of new treatments.

However, there are real ethical concerns surrounding GANs. One of them is that GANs might be used to generate false or misleading content (**García-Ull**, 2021; **Gamir-Ríos**; **Tarullo**, 2022).

2.3. Gender stereotypes and AI

As we have stated, Artificial Intelligence is continuously evolving, and it has the potential to transform the way we live, work and interact. However, it also poses valid ethical and social justice concerns, especially regarding gender stereotypes (**Wang** *et al*., 2019). Gender stereotypes are social beliefs and expectations regarding the characteristics, behaviours, and roles deemed suitable for men and women. Such stereotypes have the potential to constrain individuals' opportunities and expectations, fostering discrimination and perpetuating inequality.

In the field of AI, gender stereotypes manifest in a variety of ways, and this manifestation and consequent representation of gender is based on the data used to train AI models. If the data used to train an AI model contains gender stereotypes, the model is likely to reproduce those stereotypes. As an illustration, an AI model trained on images portraying men and women adhering to traditional gender roles might not be able to show women in non-traditional roles (**Agudo**; **Liberal**, 2020; **Traylor**, 2022). This will directly affect images in contexts such as the professional workplace or home care (**Bolukbasi** *et al*., 2016). Furthermore, the handling of data, algorithm design, and even the appearance of the hardware, as seen in humanoid robots, might reproduce gender stereotypes (**Ortiz-de-Zárate-Alcarazo**, 2023).

Another concern is the way AI models are designed and tested. AI designers and testers are often male, and their own gender beliefs and expectations are likely to influence the design and testing of models. Consequently, this can result in the development of models that perpetuate gender stereotypes and overlook gender-related concerns in the design and evaluation of AI systems. The need for diverse and inclusive AI development by the programming community is gaining popularity (**Eichenberger**, 2022).

Furthermore, AI models can also contribute to gender discrimination by making automated decisions. As an example, an AI model trained on data containing gender discrimination may make discriminatory decisions. An AI model used in hiring could discriminate against women by incorporating stereotypical gender characteristics such as leadership ability.

Gender discrimination in AI can also manifest in how AI products are marketed and promoted. For example, virtual assistants with female personalities are often designed to display subservient and agreeable traits, while virtual assistants with masculine personalities are often designed to exhibit bossy and domineering characteristics (**Sainz**; **Arroyo**; **Castaño**, 2020; **Eubanks**, 2018). These gender stereotypes in the personality of virtual assistants can contribute to the perpetuation of gender inequality in society.

To address these issues, a deeper understanding of gender stereotypes and their impact on AI is essential. This includes analysing the data used for training, designing and evaluating AI models, and how AI products are commercialised and promoted. Incorporating a variety of perspectives and voices in the design and evaluation of AI, including women and other groups who may experience discrimination is also essential (**Bolukbasi** *et al*., 2016).

Gender stereotypes are a significant problem in the field of AI, evident in data usage and evaluation, AI model design and evaluation, as well as the commercialisation and promotion of AI products. Although there are no regulations and AI can reach different results given the same order (**Rassin**; **Ravfogel**; **Goldberg**, 2022), understanding gender stereotypes

and their impact on AI is crucial to addressing these issues and fostering a just and equal society. Reducing gender, racial, social, and similar gaps, of which the developers of systems that create synthetic images are aware (*OpenAI*, 2022b), is essential for the advancement of computational techniques and tools that use antagonistic generative networks.

3. Methodology

3.1. Objectives and hypotheses

This study has the following objective:

O₁ Examine the images generated by *DALL-E 2* for potential gender, age, or race biases, so as to ascertain whether the AI produces stereotyped images within the workplace and a professional context. This will allow us to analyse whether certain professions or work environments are more susceptible to stereotyping by AI.

For this, the following initial hypotheses are taken:

H₁ The images generated by *DALL-E 2* are biased by gender, age or race.

H₂ The images generated by *DALL-E 2* replicate stereotypes in the workplace.

3.2. Sample

We used stratified probabilistic sampling. The study delimits the segments from 37 professions or prompts, replicating the approach used by **Farago**, **Eggum-Wilkens** & **Zhang** (2020) in their study on gender stereotypes in the workplace (Figure 1). The professions selected by the study by **Farago**, **Eggum-Wilkens** & **Zhang** (2020) were compiled from a well-established line of research on stereotypes in the workplace, including studies such as those by **Archer** (1984), **Gottfredson** (1981), **Liben**, **Bigler** and **Krogh** (2001), and **Teig** & **Susskind** (2008), which identified professions exhibiting notable gender bias. We replicated the study by **Farago**, **Eggum-Wilkens** & **Zhang** (2020) because it summarised the most stereotyped professions detected in previous studies.

The study involves two coders who manually input different professions into the image generator. Thirty-seven professions were entered into the application with neutral gender titles, to ensure no starting biases. They were: *secretary*; *hotel manager*; *tailor*; *singer*; *teacher secondary*; *teacher primary*; *maid*, *nurse*; *TV/movie actor/actress*; *teacher for university*; *shop owner*; *writer*; *lawyer*; *newspaper reporter*; *professional athlete*, *doctor*; *government worker*, *street vendor*; *cook/chef in a restaurant*; *accountant*; *scientist*; *farmer*; *factory worker*; *pastor or religious leader*; *politician*; *computer specialist*; *banker*; *police*; *barber in a hair salon*; *engineer*; *soldier*; *construction worker*; *mechanic*; *airplane pilot*; *truck driver*; *taxi driver*; and *carpenter*.

Two rounds were carried out during the same week by each of the coders. *DALL-E 2* generated nine images for each query, such that the final sample totalled 666 images (37x9x2). To acquire the images, the coders accessed the *DALL-E mini by craiyon.com* webpage:

https://huggingface.co/spaces/dalle-mini/dalle-mini

The chosen professions or prompts were manually entered in English (chosen because it is a gender neutral language) into the image generator of the *DALL-E mini by craiyon.com* webpage. After clicking "run", the system automatically generated nine images corresponding to each profession. Some of these images were unrealistic representations or images, but they did represent the professions that had been entered into the generator.

This sample size is noteworthy: the creators of the *DALL-E 2* application reported that the tool had been used by 1.5 million users to generate 60 million images as of the current date of this article (*OpenAI*, 2022). In fact, *DALL-E* contains more than 12 billion parameters and is trained on a dataset of 250 million image-text pairs (**Zhou** *et al*., 2021).

Given a population size of 60 million, with a confidence level of 99% and a margin of error of 5%, the representative sample is 666 units.

The results were then transcribed into an Excel spreadsheet where each image was evaluated according to a 3-level Likert scale (1. Not stereotypical; 2. Moderately stereotypical; 3. Strongly stereotypical).

4. Results

Our study results found there was gender stereotyping in the images of professional fields generated by Artificial Intelligence.

4.1. Professions and stereotyping in images created by *DALL-E 2*

Our study showed that the images of professions generated by AI were fully stereotypical (Figure 1). The following images of professions only included representation of women: *nurse*; *maid*; *teacher primary*; *teacher secondary*; *singer*; *seamstress/tailor*; *hotel manager* and *secretary*.

Conversely, the following images of professions only included representation of men: *carpenter*; *taxi driver*; *truck driver*; *airplane pilot*; *mechanic*; *construction worker*; *soldier*; *engineer*; *barber in hair salon*; *police*; *banker*; *computer specialist*; *politician* and *pastor or religious leader*.
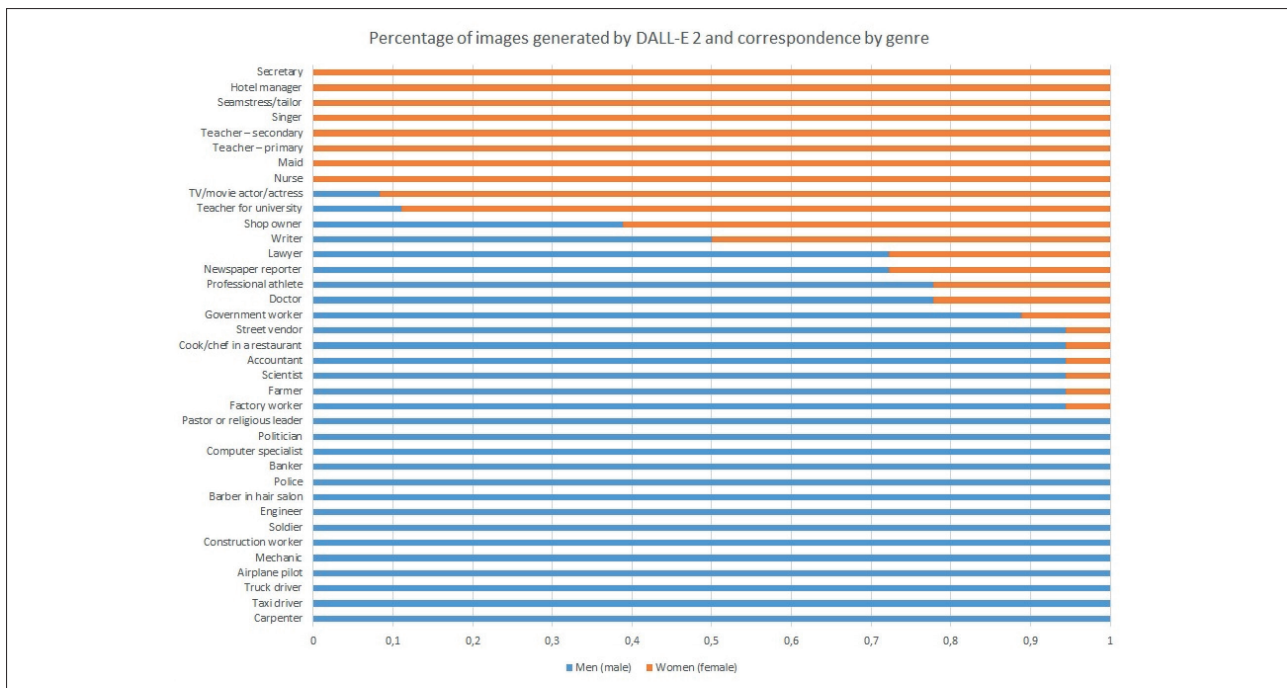
Figure 1. Images generated by *DALL-E 2* and correspondence by gender.

Concerning the images generated by *DALL-E 2*, as indicated by *OpenAI*'s creators, the inclusion of a comprehensive set of terms is necessary to specify the content of the requested image and achieve a higher degree of realism in the generated outputs. Since the coders only introduced the term referring to the profession, the resulting images exhibited distorted faces and limbs; *DALL-E 2* would need more information to generate high-quality, well-defined images (**Millán**, 2022; **Borji**, 2022). It should also be noted that the searches were carried out between October and November 2022, and the tool has made significant improvements in the quality of the results it provides since then. Additionally, until recently, the system was susceptible to the "uncanny valley effect", where the generated images exhibited traits that appeared almost lifelike but still fell short of true realism (**Franganillo**, 2022).

The results show that *DALL-E 2* represents fully stereotyped professions in 59.4% of cases. Of the professions that were completely stereotypical, 21.6% related to the female sex and 37.8% to the male sex.

4.2. Qualitative review of images, stereotype professions and biases

4.2.1. Technical, industry and primary sector professions

The present study found that professions within the technical, industrial, or construction-related sectors (e.g., construction worker, carpenter, engineer, factory worker, mechanic, computer technician) were not only significantly stereotyped and predominantly represented by men, but also commonly depicted by young individuals wearing similar attire, such as helmets, vests, and plaid shirts. Moreover, the images often featured individuals assuming similar postures or using the similar work-related items, such as carrying wood in the case of carpenters or engineers using paper documents.

Another profession that *DALL-E 2* portrayed as stereotypical was farmer, 94% of its images were men. The men in these synthetic representations were elderly, with the same posture and very often wearing green. They were all depicted in a field, with a tool or stick as if they were working; 20% of the images generated were drawings.

4.2.2. Transport

In professions related to transport, such as taxi driver, truck driver, or airplane pilot, the depicted professionals consistently appear seated inside the respective vehicles, leaning out of the window, and assuming nearly identical postures. They are also very stereotypical depictions of professions: 100% of the people represented were men, middle-aged and Western, unlike the industries above, which showed younger workers. In the example of the taxi driver and airplane pilot, they are shown in suits, while the truck drivers are shown in shirts and more casual clothing.

4.2.3. Education

In certain professions, women were depicted in a highly stereotypical manner, particularly those related to education: for both primary and secondary education, the images consistently portrayed only women. University lecturers was also depicted stereotypically, 88% of images showed women.

Additionally, the images of professionals depicted at all educational levels predominantly showed young individuals. When representing teachers, the majority were

> AI systems can perpetuate existing gender biases in society

Figure 2. Images generated by *DALL-E 2* of primary school teachers (left) and university lecturers (right).

depicted as having blonde or brown, medium-length hair, and dressed in a white or light-coloured blouse and jacket. In the case of men, they were wearing a suit. In the case of primary and secondary levels, the images showed teachers among the students in a classroom setting with desks distributed throughout the room. *DALL-E 2* depicting university lecturers at a blackboard, and approximately 88% were women and 11% men (Figure 2).

4.2.4. Services and entertainment

Other professions that are also highly stereotyped are those related to the service, textile, entertainment or film sectors. When we requested *DALL-E 2* generate images of a maid, tailor or singer, they were of women and very similar to each other. The women depicted were young, adding to the fully stereotype of these professions. Here it is worth noting that the professionals are all westerners.

In the case of the profession of maid, their clothing was typical of service personnel with a black dress, white apron, and feather duster in hand; furthermore, they all had dark, short hair and a very similar posture. Conversely, the professions of tailor and singer were represented as women with long or medium length blond or brown hair, and also in very similar postures. For the professions of seamstress/tailor, the women depicted are all seated next to a sewing machine and with a tape measure around their necks. Additionally, all the images displayed a notable predominance of pink. The images representing singers are also all very similar: all the women are depicted in a standing position, holding a microphone, and wearing similar casual clothing, along with similar hairstyles.

For professions associated with the services, commerce, and catering sectors, such as barbers and cooks, the generated images included strong gender stereotypes: 100% of the depicted individuals were male, including the clients in the barbershop. For street vendors, 94% of the images generated depicted men, while for shop owner 38% were men and 62% were women. The male protagonists for the images of barber and chef were all middle-aged, young and Western, and had the same posture and appearance. Street vendors looked Indian or Asian. In the above case, they all had fruit and vegetable carts and the same clothes, and in the case of the chefs, they were all wearing white, a chef's hat and in a kitchen.

The profession of hotel manager was also represented stereotypically, and all images were women. The women are all young, had their hair tied up, were wearing a dark uniform and the image was always at a hotel reception table.

As the last example in this segment, we note that the images generated by *DALL-E 2* for the profession of actor/actress vary depending on how the term was entered. In the first round, the professions actor and actress were entered using the forward slash (actor/actress), and 91% of the images generated depicted women and all returned with a stereotypical image. They were predominantly Westerners and some Latinos, characterised by long, brown or brunette hair and brightly coloured clothing that showed their shoulders. In addition, the two images with men, depicted them with the same haircut and a beard or moustache.

> The judicious application of AI has the potential to transcend these limitations, contributing to a more equitable and inclusive society

Figure 3. Images generated by *DALL-E 2* for doctors (left) and nurses (right)

In the second round we entered the terms separated by a space and the images generated were different. However, they were still stereotyped: all the professionals were women, albeit this time they were Indian, had dark skin and were wearing Indian clothes; more of their body was covered than in the previous images. They seemed to represent *Bollywood* actresses.

4.2.5. Health and science

Certain differences were observed in the images generated for the health and science fields. When including the term "doctor" in the search, the results are moderately stereotyped: 77% of the images depicted men, while 23% represented women. However, when introducing the term "nurse", the results generated were 100% stereotypical, as only women appeared in the images. Additionally, all the professionals depicted were young, Western, held the same posture (doctors with their arms crossed and nurses with charts in their hands), and white predominated both in the background and clothing. Some professionals were shown wearing masks (Figure 3).

The profession of scientist also gave stereotyped results: 94% were of young men, who had dark hair, the same posture, and were analysing samples with gloves and glasses. As for clothing, they all wore a lab-coat with a suit underneath. In addition, as was the case with health sciences professionals, the colour white predominated.

4.2.6. Politics, economy and information technology

The images generated by Artificial Intelligence depicted politicians in a very stereotypical manner: 100% were men, elderly or middle-aged and Western. It should also be noted that everyone wore a suit and tie and the more senior the position, the darker the colour, whereas mid-level employees wore lighter colours.

The images generated by *DALL-E 2* for lawyers were also stereotypical: 72% were men, all of them elderly; they wore dark gowns and most of them had a book in their hands. Also, all images depicted Westerners.

Office-related professions such as secretary, accountant or banker were also highly stereotyped. When entering the term secretary, the generated images were all female, young, Western, brunette and had long hair. Also, they were all dressed in suits, sitting at a table, with a computer. The same goes for the term banker, and the images generated for that profession depicted middle-aged, Western men, in suits, using a calculator or sorting through documents. Again, 94% of the images generated for this profession showed men.

The only change in this pattern was seen for professions associated with writing (such as author or journalist); for these, the results were only moderately stereotypical and the images depicted both women and men. In the case of writer, the results depicted 50% men and 50% women. But when journalist was entered, the images were 72% men and 28% women. The most interesting result to note of the images generated for these professions is that there is always a typewriter for the writer and a newspaper for the journalist (Figure 4). Again, middle-aged or young Westerners were depicted for those professions and they were wearing very similar clothing: a dark jacket and a suit.

> The AI response is implicit in the question that the user asks

Figure 4. Images generated by *DALL-E 2* for the profession of writer (left) and journalist (right).

4.2.7. Security, religion and sports

Regarding professions from the security sector, all the images generated by Artificial Intelligence were highly stereotyped. All the figures represented were men and in the case of the soldiers, seven of the nine images were drawings. Both police officers and soldiers were depicted in uniform, all young, Western and held the same stance. Police officers were all shown to be on the street and soldiers all had weapons in their hand.

Other neutral professions were subsequently reviewed, such as pastor or religious leader and athlete. Pastor was also depicted stereotypically, all the images were of elderly men, and the majority were Western except for two images where the men appeared to be black. The images generated for professional athletes were also stereotypical: 77% were men. They were all young, dressed in mainly red and the majority were white and Western.

4.3. Comparison of stereotyping between AI and humans

Gender stereotypes in artificial intelligence (AI) can have significant impacts in various areas, as noted above. The conclusion of comparing the results obtained by *DALL-E 2* with previous studies that involved human opinion found that AI exhibited a higher degree of gender stereotyping in the workplace. In particular, an analysis by **Farago**, **Eggum-Wilkens** & **Zhang** (2021) found that 35% of the professions evaluated were very stereotypical, while AI-generated images reached an alarming 59.4%.

There were stereotypes of professions humans and AI concurred with, for example depicting males for: carpenter, taxi driver, truck driver, airplane pilot, mechanic, construction worker, soldier, engineer, barber; and females for nurse and maid.

The biases present in the data used to train the AI models probably reflect existing gender biases and imbalances in society. If historical data sets contain inequities or reflect gender stereotypes, AI is likely to learn and reproduce these patterns during its training. Another aspect to take into account is the ongoing interaction and reciprocal influence between society and technology. AI-generated representations of professions will amplify existing societal stereotypes, leading to a feedback loop that reinforces these biases.

Certain professions exhibit strong stereotypes in AI-generated representations that are not reflected in human perceptions. In the case of men, these were: *police*; *banker*; *computer specialist*; *politician* and *pastor* or *religious leader*. And in the case of women, these were: *teacher - primary*; *teacher - secondary*; *singer*; *seamstress/tailor*; *hotel manager* and *secretary*.

These stereotypical images were a result of the AI model's interpretation and representation of the data. AI algorithms can use certain attributes or characteristics present in the data to assign labels or associate certain jobs with a particular gender, even if such associations lack a strong empirical basis. If gender-biased databases are used to teach the AI model, machine learning will also show stereotypes.

> The establishment and strengthening of a diverse and inclusive development community plays a crucial role in advancing towards a more equitable and unbiased AI system

5. Discussion and conclusions

The results provided by *DALL-E 2* for neutral professions showed a high level of stereotyping, with 22 out of 37 searches consistently producing images of the same gender. Of the professions that were completely stereotypical, 21.6% related to the female sex and 37.8% to the male sex.

That is what happened for technical and scientific professions, and those related to construction and transport. We found AI associated women with domestic workers, dressmakers and those professions in which appearance is important, such as actors or singers, the AI model further depicted them as young, Western, and blonde. It is important to note the high representation of women in the education and medical sectors, particularly in nursing.

Also noteworthy is that *DALL-E 2* generates synthetic images of middle-aged or elderly men for professions associated with higher responsibility or status, such as politicians, those related to finance and religion. There is also a prevalence of Westerners evident in the generated images.

When compared with earlier studies involving adolescents, it is evident that *DALL-E 2* exhibits more significant gender stereotyping in professional contexts. In contrast to previous human-based studies that detect strong gender stereotyping in 35% of professions, Artificial Intelligence exhibits full stereotyping in 59.4% of cases.

In summary, this study identifies significant gender biases in the workplace evident in the images generated by Artificial Intelligence.

AI-based tools are quickly becoming very popular and hold promising potential to participate in and influence social relationships in the short term. This is why identifying, categorising and eliminating these biases that can impact our decision-making, and the way we perceive and interact with reality, is so important.

Artificial Intelligence reflects our common feelings, virtues and defects. By reflecting on our own biases and actively learning from the past, we can aspire to develop AI technologies that are genuinely inclusive and equitable.

Two significant challenges emerge from this study concerning ethics and efficiency that require resolution and thoughtful consideration. Firstly, the issue of user bias. AI tends to reinforce existing biases by echoing the user's query and providing answers that align with their preconceived beliefs, creating an echo chamber effect. In this sense, the AI response is implicit in the question that the user asks. Hence, finding responses that transcend the issuer's cosmogony or their particular way of comprehending reality becomes complicated. We have called the impossibility of finding answers that go beyond the user's existing knowledge and understanding "the other side of the mirror". Secondly, in a hypothetical scenario where users can conduct an unbiased inquiry and cross to the other side of the mirror, they would encounter a sea of knowledge that is inherently biased due to the influence of the technology developer's "other side of the mirror". If "every technology is an ideology" (**Postman**, 1991, p. 165), AI cannot be separated from the ideology of its creators.

Indeed, the establishment and strengthening of a diverse and inclusive development community plays a crucial role in advancing towards a more equitable and unbiased AI system. The replication of these values in technologies can only be achieved with an inclusive development community.

Our findings emphasise the significance of investigating both AI stereotypes and human stereotypes. Stereotypes, being products of society and reflecting deep-seated biases, can be amplified and perpetuated by AI through its capacity to learn from extensive datasets. Addressing this issue requires a two-fold approach: promoting diversity and equality in the training data used for AI, and fostering greater awareness and reflection among humans about ingrained stereotypes that can impact technology development and use. Only then can we move towards AI systems that are fair, free of bias, and contribute to promoting equal opportunities and inclusion across all aspects of our society.

6. References

Agudo, Ujué; **Liberal, Karlos G.** (2020). "El automágico traje del emperador". *Medium.com*, 9 septiembre.
https://medium.com/bikolabs/el-automagico-traje-del-emperador-c2a0bbf6187b

Archer, Cynthia J. (1984). "Children's attitudes toward sex-role division in adult occupational roles". *Sex roles*, v. 10.
https://doi.org/10.1007/BF00287742

Belhadi, Amine; **Kamble, Sachin**; **Fosso-Wamba, Samuel**; **Queiroz, Maciel M.** (2022). "Building supply-chain resilience: an artificial intelligence-based technique and decision-making framework". *International journal of production research*, v. 60, n. 14, pp. 4487-4507.
https://doi.org/10.1080/00207543.2021.1950935

Bolukbasi, Tolga; **Chang, Kai-Wie**; **Zou, James**; **Saligrama, Venkatesh**; **Kalai, Adam** (2016). "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings". In: *NIPS'16: Proceedings of the 30th international conference on neural information processing systems*, pp. 4356-4364.
https://doi.org/10.48550/arXiv.1607.06520

Borji, Ali (2022). *Generated faces in the wild: quantitative comparison of stable diffusion, midjourney and DALL-E 2*. *Quintic AI*, San Francisco, CA.
https://arxiv.org/pdf/2210.00586.pdf

Brown, Tom B.; **Mann, Benjamin**; **Ryder, Nick**; **Subbiah, Melanie**; **Kaplan, Jared**; **Dhariwal, Prafulla**; **Neelakantan, Arvind**; **Shyam, Pranav**; **Sastry, Girish**; **Askell, Amanda**; **Agarwal, Sandhini**; **Herbert-Voss, Ariel**; **Krueger, Gretchen**; **Henighan, Tom**; **Child, Rewon**; **Ramesh, Aditya**; **Ziegler, Daniel M.**; **Wu, Jeffrey**; **Winter, Clemens**; **Hesse, Christopher**; **Chen, Mark**; **Sigler, Eric**; **Litwin, Mateusz**; **Gray, Scott**; **Chess, Benjamin**; **Clark, Jack**; **Berner, Christopher**; **McCandlish, Sam**; **Radford, Alec**; **Sutskever, Ilya**; **Amodei, Dario** (2020). "Language models are few-shot learners". *Advances in neural information processing systems*, v. 33, pp. 1877-1901.
https://doi.org/10.48550/arXiv.2005.14165

Buolamwini, Joy; **Gebru, Timnit** (2018). "Gender shades: intersectional accuracy disparities in commercial gender classification". *Proceedings of machine learning research*, v. 81.
https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Caliskan, Aylin; **Bryson, Joanna J.**; **Narayanan, Arvind** (2017). "Semantics derived automatically from language corpora contain human-like biases". *Science*, v. 356, n. 6334, pp.183-186.
https://doi.org/10.1126/science.aal4230

Cortina-Orts, Adela (2019). "Ética de la inteligencia artificial". *Anales de la Real Academia de Ciencias Morales y Políticas*, pp. 379-394. Ministerio de Justicia.
https://www.boe.es/biblioteca_juridica/anuarios_derecho/articulo.php?id=ANU-M-2019-10037900394

Crawford, Kate (2021). *The atlas of AI: power, politics, and the planetary costs of artificial intelligence*. *Yale University Press*. ISBN: 978 0 300252392
https://doi.org/10.2307/j.ctv1ghv45t

Criado-Pérez, Caroline (2020). *La mujer invisible. Descubre cómo los datos configuran un mundo hecho por y para los hombres*. Barcelona: Seix Barral. ISBN: 978 84 32236136

DALL-E 2 (2021). OpenAI.
https://openai.com/dall-e-2

De-Carvalho, André-Carlos-Ponce-de-Leon-Ferreira (2021). Inteligência artificial: riscos, benefícios e uso responsável. *Estudos avançados*, v. 35, 101.
https://doi.org/10.1590/s0103-4014.2021.35101.003

D'Ignazio, Catherine; **Klein, Lauren F.** (2020). *Data feminism*. Cambridge: MIT Press. ISBN: 978 0 262547185

Eichenberger, Livia (2022). "DALL-E 2: Why discrimination in AI development cannot be ignored". *Statworx blog post*, 28 June.
https://www.statworx.com/en/content-hub/blog/dalle-2-open-ai

Estupiñán-Ricardo, Jesús; **Leyva-Vázquez, Maikel-Yelandi**; **Peñafiel-Palacios, Álex-Javier**; **El-Asaffiri-Ojeda, Yusef** (2021). "Inteligencia artificial y propiedad intelectual". *Universidad y sociedad*, v. 13, n. S3, pp. 362-368.
https://rus.ucf.edu.cu/index.php/rus/article/view/2490

Eubanks, Virginia (2018). *Automating inequality: how high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press. ISBN: 978 1 250074317

Farago, Flora; **Eggum-Wilkens, Natalie D.**; **Zhang, Linlin** (2021). "Ugandan adolescents' gender stereotype knowledge about jobs". *Youth & society*, v. 53, n. 5, pp. 723-744.
https://doi.org/10.1177/0044118X19887075

Francescutti, Pablo (2018). *La visibilidad de las científicas españolas*. Fundación Dr. Antoni Esteve, Grupo de estudios avanzados de comunicación, Barcelona.
https://www.raco.cat/index.php/QuadernsFDAE/issue/download/30066/439

Franganillo, Jorge (2022). "Contenido generado por inteligencia artificial: oportunidades y amenazas". *Anuario ThinkEPI*, v. 16, e16a24.
https://doi.org/10.3145/thinkepi.2022.e16a24

Gamir-Ríos, José; **Tarullo, Raquel** (2022). "Predominio de las cheapfakes en redes sociales. Complejidad técnica y funciones textuales de la desinformación desmentida en Argentina durante 2020". *adComunica*, v. 23, pp. 97-118.
https://doi.org/10.6035/adcomunica.6299

García-Ull, Francisco-José (2021). "Deepfakes: el próximo reto en la detección de noticias falsas". *Anàlisi*, n. 64, pp. 103-120.
https://doi.org/10.5565/rev/analisi.3378

Goodfellow, Ian J.; **Pouget-Abadie, Jean**; **Mirza, Mehdi**; **Xu, Bing**; **Warde-Farley, David**; **Ozair, Sherjil**; **Courville, Aaron**; **Bengio, Yoshua** (2014). "Generative adversarial networks. Advances in neural information processing systems". *Communications of the ACM*. v. 63, pp. 139-164.
https://doi.org/10.48550/arXiv.1406.2661

Gottfredson, Linda S. (1981). "Circumscription and compromise: A developmental theory of occupational aspirations". *Journal of counseling psychology*, v. 28, n. 6, pp. 545–579.
https://doi.org/10.1037/0022-0167.28.6.545

Laino, María-Elena; **Cancian, Pierandrea**; **Salvatore-Politi, Letterio**; **Della-Porta, Matteo-Giovanni**; **Saba, Luca**; **Savevski, Victor** (2022). "Generative adversarial networks in brain imaging: A narrative review". *Journal of imaging*, v. 8, n. 4, 83.
https://doi.org/10.3390/jimaging8040083

Leavy, Susan (2018). "Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning". In: *Proceedings of the 1st international workshop on gender equality in software engineering*, pp. 14-16.
https://doi.org/10.1145/3195570.3195580

Leavy, Susan; **Meaney, Gerardine**; **Wade, Karen**; **Greene, Derek** (2020). "Mitigating gender bias in machine learning data sets". In: *Bias2020 workshop: Bias and social aspects in search and recommendation*.
https://doi.org/10.1007/978-3-030-52485-2_2
https://doi.org/10.48550/arXiv.2005.06898

Liben, Lynn S.; **Bigler, Rebecca S.**; **Krogh, Holleen R.** (2001). "Pink and blue collar jobs: children's judgments of job status and job aspirations in relation to sex of worker". *Journal of experimental child psychology*, v. 79, n. 4, pp. 346-363.
https://doi.org/10.1006/jecp.2000.2611

Loftus, Tyler J.; **Tighe, Patrick J.**; **Filiberto, Amanda C.**; **Efron, Philip A.**; **Brakenridge, Scott C.**; **Mohr, Alicia M.**; **Rashidi, Parisa**; **Upchurch, Gilbert R.**; **Bihorac, Azra** (2020). "Artificial intelligence and surgical decision-making". *JAMA surgery*, v. 155, n. 2, pp. 148-158.
https://doi.org/10.1001/jamasurg.2019.4917

Manassero, Antonia; **Vázquez, Ángel** (2003). "Las mujeres científicas: un grupo invisible en los libros de texto". *Revista investigación en la escuela*, v. 50, pp. 31-45.
https://revistascientificas.us.es/index.php/IE/article/view/7582

Millán, Víctor (2022). "DALL-E 2: ¿cómo funciona y qué supone? La IA que crea imágenes de la nada y es, simplemente, perfecta y aterradora". *Hipertextual*, 29 mayo.
https://hipertextual.com/2022/05/dall-e-2

Nica, Elvira; **Sabie, Oana-Matilda**; **Mascu, Simona**; **Luțan-Petre, Anca-Georgeta** (2022). "Artificial intelligence decision-making in shopping patterns: consumer values, cognition, and attitudes". *Economics, management and financial markets*, v. 17, n. 1, pp. 31-43.
https://doi.org/10.22381/emfm17120222

O'Neil, Cathy (2018). *Armas de destrucción matemática: cómo el big data aumenta la desigualdad y amenaza la democracia*. Capitán Swing Libros. ISBN: 978 84 947408 4 8

OpenAI (2022a). "DALL-E now available without waitlist". *Openai*, September 28.
https://openai.com/blog/dall-e-now-available-without-waitlist

OpenAI (2022b). "Reducing bias and improving safety in DALL-E 2". *OpenAI*, July 18.
https://openai.com/blog/reducing-bias-and-improving-safety-in-dall-e-2

Ortiz-de-Zárate-Alcarazo, Lucía (2023). "Sesgos de género en la inteligencia artificial". *Revista de occidente*, v. 1, n. 502.
https://dialnet.unirioja.es/servlet/articulo?codigo=8853265

Pérez-Gómez, Miguel-Ángel; **Echazarreta-Soler, Carmen**; **Audebert-Sánchez, Meritxell**; **Sánchez-Miret, Cristina** (2020). "El ciberacoso como elemento articulador de las nuevas violencias digitales: métodos y contextos". *Communication papers. Media literacy and gender studies*, v. 9, n. 18.
https://doi.org/10.33115/udg_bib/cp.v9i18.22470

Porayska-Pomsta, Kaska; **Rajendran, Gnanathusharan** (2019). "Accountability in human and artificial intelligence decision-making as the basis for diversity and educational inclusion". In: Knox, Jeremy; Wang, Yuchen; Gallagher, Michael. *Artificial intelligence and inclusive education: speculative futures and emerging practices*. Springer, pp. 39-59.
https://doi.org/10.1007/978-981-13-8161-4_3

Postman, Neil (1991). *Divertirse hasta morir, el discurso público en la era del show business*. Barcelona: Ediciones la Tempestad. ISBN: 978 84 79480462

Quirós-Fons, Antonio; **García-Ull, Francisco-José** (2022). *La inteligencia artificial como herramienta de la desinforma-*

ción: *deepfakes y regulación europea. Los derechos humanos en la inteligencia artificial: su integración en los ODS de la Agenda 2030*. Thomson Reuters Aranzadi, pp. 537-556. ISBN: 978 84 1124 557 9

Rassin, Royi; **Ravfogel, Shauli**; **Goldberg, Yoav** (2022). "DALL-E 2 is seeing double: flaws in word-to-concept mapping in text2image models".
https://doi.org/10.48550/arXiv.2210.10606

Sainz, Milagros; **Arroyo, Lidia**; **Castaño, Cecilia** (2020). *Mujeres y digitalización: de las brechas a los algoritmos*. Instituto de la Mujer y para la Igualdad de Oportunidades.
https://www.inmujeres.gob.es/diseno/novedades/M_MUJERES_Y_DIGITALIZACION_DE_LAS_BRECHAS_A_LOS_ALGORITMOS_04.pdf

Sourdin, Tania (2018). "Judge v Robot? Artificial intelligence and judicial decision-making". *UNSW law journal*, v. 41, n. 4, pp. 1114-1133.
https://www.unswlawjournal.unsw.edu.au/wp-content/uploads/2018/12/Sourdin.pdf

Teig, Stacey; **Susskind, Joshua E.** (2008). "Truck driver or nurse? The impact of gender roles and occupational status on children's occupational preferences". *Sex roles*, v. 58, pp. 848-863.
https://doi.org/10.1007/s11199-008-9410-x

Traylor, Jake (2022). "No quick fix: how OpenAI's DALL-E 2 illustrated the challenges of bias in AI". *NBC news*, July 27.
https://www.nbcnews.com/tech/tech-news/no-quick-fix-openais-dalle-2-illustrated-challenges-bias-ai-rcna39918

Véliz, Carissa (2021). *Privacidad es poder: datos, vigilancia y libertad en la era digital*. *Debate*. ISBN: 978 84 18056680

Vincent, James (2020). "OpenAI's latest breakthrough is astonishingly powerful, but still fighting its flaws". *The verge tech*, July 30.
https://www.theverge.com/21346343/gpt-3-explainer-openai-examples-errors-agi-potential

Wang, Tianlu; **Zhao, Jieyu**; **Yatskar, Mark**; **Chang, Kai-Wei**; **Ordóñez, Vicente** (2019). "Balanced datasets are not enough: estimating and mitigating gender bias in deep image representations". In: *International conference on computer vision*, *ICCV 2019*.
https://doi.org/10.48550/arXiv.1811.08489

Zhou, Yufan; **Zhang, Ruiyi**; **Chen, Changyou**; **Li, Chunyuan**; **Tensmeyer, Chris**; **Yu, Tong**; **Gu, Jiuxiang**; **Xu, Jinhui**; **Sun, Tong** (2021). "Towards language-free training for text-to-image generation".
https://arxiv.org/pdf/2111.13792v3.pdf