Designed to exclude? Investigating visual ageism and sexism in text-to-image AI generative models

Gianluca De Ninno[1,2,*,†], Beatrice Melis[1,2,†] and Francesca Belotti[3,†]

[1]*Computer Science Area, Gran Sasso Science Institute: L'Aquila, Italy*
[2]*Department of Computer Science, University of Pisa: Pisa, Italy*
[3]*Department of Humanities, University of L'Aquila: L'Aquila, Italy*

Abstract

In ongoing debate on the biases embedded in generative artificial intelligence (GenAI) systems, this paper illustrates the results of qualitative research unveiling the sexist and ageist stereotypes conveyed through images generated by ChatGPT-4o. The theoretical framework draws upon previous studies on digital ageism and sexism, focusing on visualities as carriers of socio-cultural repertoires. Accordingly, we approach GenAI as socio-technical systems and communicative agents, having a human cultural matrix. On a methodological level, we experimented with an incipient research protocol structured around three conversational threads, allowing for the contestual analysis of visual outputs across different interaction scenarios in controlled settings. The prompts refer to daily activities that involve the use of different technologies. Preliminary findings from the critical visual analysis reveal an *overrepresentation* of male and young people and, conversely, an *invisibilization* of marginalized social groups, such as female (and) older people and gender non-conforming individuals.

Keywords

Text-to-image, Visual ageism, Gender bias, Generative AI, Digital ageism, Digital sexism

1. Introduction

In recent years, text-to-image (T2I) AI generative models have become increasingly high-performing, expanding the possibilities of digital content creation. However, these models have also raised ethical concerns about biases that appear embedded in their outputs [1, 2]. While increasing attention has been reserved for gender-based ones [3, 4, 5], other forms of discrimination are underexplored, especially biases based on age that mostly affect older people (i.e., "ageism"). These latter, can also enhance specific stereotypes and forms of marginalization against older women when combined with gender-based biases [6, 7]. Considering that images are capable of stimulating the creation of meaning [8], several scholars address with concern the current and future spread of synthetically created images in the digital landscape [9, 10], including their use on social media platforms and in news outlets. Non-expert users might not be able to distinguish a real image from a synthetic one, with implications for the shaping of a biased cultural imaginary [11]. Thus, it is desirable for AI-driven technologies to be able to portray accurate images of the social world and fairly represent the diversity of human experiences [12].

In our study, we critically analyze synthetic images generated by a GenAI model (namely, ChatGPT-4omni) in order to unveil possible sexist and ageist representations, considering them as "mirrors" of socio-technical shortcomings. In this, we approach the field of T2I from both a socio-cultural and technical perspective. On one hand, we consider GenAI systems as communicative agents mimicking human sociality [13, 14, 15] that have been socialized to the socio-cultural repertoires of the environment in which they are developed and deployed [16, 14, 17], and can therefore reproduce discriminatory attitudes and behaviors [18, 19], for example, towards women, older people and LGBTQIA+ individuals.

HHAI-WS 2025: Workshops at the Fourth International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 9-13, 2025, Pisa, Italy
*Corresponding author.
†These authors contributed equally.
gianluca.deninno@gssi.it (Gianluca De Ninno); beatrice.melis@gssi.it (Beatrice Melis); francesca.belotti@univaq.it (Francesca Belotti)

On the other hand, we look at GenAI systems as autonomous technological products composed of technical components (i.e., algorithms, data and infrastructure) that, like an atlas, map the environmental, economic, cultural, and geopolitical resources and dynamics involved in them [6]. GenAI models in general and specifically T2I systems are powered by machine learning techniques - that is, statistical prediction methods that learn patterns associating textual descriptions with corresponding visual representations, based on their co-occurrence in training datasets. As argued by Salvaggio (2023: [20]), "AI images are data patterns inscribed into pictures, and they tell us stories about these image-text datasets and the human decisions behind them." Therefore, AI-generated images convey embedded narratives and knowledge structures, leaving users (and mostly, *non*-expert users) unintentionally exposed to potential harm [21].

2. Background and rationale

Why synthetic images. T2I models represent a groundbreaking change in AI-driven content creation, enabling users and professionals to generate complex and detailed images from textual instructions (i.e., prompts). Just like traditional photographic or illustrative processes, T2I models might carry socio-cultural biases [20], which might then be reflected and amplified by their visual outputs [22].

Vision is featured as something that has a cultural centrality, whereby images are able to embody and express cultural values and societal norms [8]. Visual media, such as photographs, illustrations, and films, are not only reflections of reality, but are active agents in shaping how identities, power structures, and social norms are understood and reproduced [23, 24, 25]. This is why Schroeder (2006, [8]) emphasizes the so-called "politics of representation" [8]: visualities influence how identities are perceived, embodied and, hence, performed [26], even in the interactions with others within which we negotiate who we are [27]. In the domain of T2I technologies, the generated images are closely related to both the datasets that train the models and the algorithms that allow the synthesis [28]. Such technologies generate images by translating textual descriptions into visual representations. The models learn correlations between text and images using large datasets of text-image pairs, while the algorithms synthesize images that align with the provided text. In this dynamics, shortcomings and errors might occur, such as multi-modal hallucinations [29, 30], that are defined as inconsistencies between the generated textual response and the associated visual content, which are linked to the quality and quantity of training data as well as to the statistical misalignment between textual and visual modalities during the training [31]. For instance, a lack of diversity in the dataset or a loss of descriptive detail in the image captions can contribute to these mismatches, highlighting the need for critical examination of the reliability and representational accuracy of the outputs produced by such systems.

Since the training datasets are extracted from specific socio-cultural contexts, systematic errors might occur [19, 18]. This recalls the phenomenon of "bias-in, bias-out" [32], meaning that biases introduced either by AI systems or by humans can reinforce each other, in a feedback loop that ends up increasingly distorting perceptions, especially via images [33, 34].

In this regard, AI-developers define training datasets as the basement of a ground truth [35], emphasizing how machines' behavior is strictly dependent on the knowledge they have acquired from the datasets they have learnt upon. Crawford (2021, [36]) argues that "truth is less about a factual representation of an agreed-upon reality and more commonly about a jumble of images scraped from whatever online sources are available" (p. 96). This implies that the notion of truth in these systems is not an objective or neutral construct, but rather a contingent and contextually determined representation of an hegemonic reality that eventually marginalize communities and experiences whose data are not available or misregarded. Therefore, directly observable features of synthetic images might suggest things about the underlying datasets [20], thus accounting for how an AI "see" (and hence represent) the world [36].

With these premises in mind, we formulated the following Research Questions:

- RQ1: What sexist and/or ageist visual content, if any, is incorporated into synthetic images?
- RQ2: What does this tell us about the socio-technical components of GenAI systems?

Why gender-based biases. Gender-related issues constitute an important point of convergence for multiple studies concerned with the biases embedded and reproduced by AI-driven technologies [37, 38, 39]. Building upon the conception of gender as a complex socio-cultural construct [40], studies on gender and digital technologies have extensively demonstrated that these latter have been historically dominated by male-centric perspectives [36]. Scholars have argued that the tech world has long reflected and reproduced gendered power structures, often marginalizing women and gender minorities both in the workforce and in design priorities [41, 42, 43]. These dynamics are further exacerbated by the underrepresentation of women in key roles across STEM fields, particularly in AI development [44], contributing to the normalization of male-default assumptions in both software and hardware design [45]. For instance, personal assistants and smart technologies encode idealized and stereotypical notions of femininity, pointing to an actual "smart wife" archetype [46]. They are designed upon codes of care and emotional support, which are still culturally considered a feminine prerogative.

This gendered disparity has also consequences on users, whose experiences, values and bodies are encoded in (and, hence, disciplined by) technological systems. As Criado Perez (2019, [47]) highlights, digital infrastructures and technologies frequently overlook the needs of non-male users, reinforcing systemic exclusions [48]. Therefore, studying AI-generated images from a gender perspective would allow uncovering sexist prejudices and stereotypes operating along the machine learning pipeline. This phenomenon has begun to be investigated in recent years [3, 4, 49], also from a non-binary perspective [5], which is often overlooked but equally crucial [50].

This concern for sexist representations extends beyond T2I technologies, and it is rooted in long-standing patterns of visual sexism, that is, the objectification, stereotyping, or exclusion of women and LGBTQIA+ individuals in visual media, which perpetuates their societal stigmatization or marginalization [51, 52]. From advertising and branding [53] to online images [54], visual cultures have historically reinforced normative ideals of femininity and masculinity, often objectifying or disregarding women [55]. These visual logics have migrated into digital environments, giving rise to what has been described as "digital sexism" [56, 57], i.e., a "pattern of direct attacks against women's identity and ideology" ([58]: p. 1702) that are perpetuated in the digital environment.

Some recent studies have also problematized the intersection of gender-based biases with those based on ethnicity in the images and texts generated by AI systems [59, 60], but what happens when sexism intersects with (late) age is still understudied.

Why age-based biases. Building on a vision of age as a socially constructed, multi-layered concept that includes biological, psychological, socio-cultural, and economic dimensions [61, 62], several scholars from different backgrounds have been questioning the invisibility of ageism, i.e., the ideologies and practices that discriminate against (older) people based on their (old) age, especially when it comes to the digital realm [63, 64, 65, 66]. In this regard, the invisibilization and stigmatization of the older people in the datafication processes on which AI systems are based are also relevant (i.e., "data ageism", [66, 67]), as they can then be reproduced or even amplified in the outputs. However, in the broad landscape of studies on the ethical and social implications of GenAI, ageism is significantly less considered than other discriminations based gender, ethnicity, or disability; and yet, it can operate in different phases of the machine learning pipeline [68, 69]. Also, Gallistl et al. (2024: [70]) have proposed to problematize GenAI systems as a web of automatically, misleadingly datafied knowledge about older populations, which opacifies the role of humans in making the data and placing older adults into fixed categories that influence their lives.

In the field of studies on ageism in communication and information technologies, a concept that is particularly relevant to our study is that of "visual ageism", here applied to the imagery conveyed by synthetic images. Loos & Ivan (2018, [71]) define visual ageism as "the social practice of visually underrepresenting older people or misrepresenting them in a prejudiced way" (p. 164). We adapt their reflections on the media domain to the context of GenAI systems, with the same concern of witnessing a conveyed "prescription" of how to age, rather than a faithful and fair "description" of how people make sense of late life [71].

Studies on this topic are still incipient. Putland et al. (2023, [72]) and Byrne et al. (2024, [73]), for instance, highlight that AI-generated images depict older people with stereotypical signs of degeneration and frailty when it comes to represent dementia and aged care nurses. Linares-Lanzman & Rosales (2024, [74]), instead, unveil specific representational patterns of both older and young people in AI-generated images, with the former appearing seated, tired, and/or frustrated, and the latter standing, active, and smiling. These authors also introduce the concept of "generative ageism" as a form of ageism that is produced and reinforced in the environment of generative AI, taking the form of text, images or videos, and intersecting with stereotypes based on age, gender and race [74].

Since "the study of the history of images of ageing is the study of the history of our ideas about ageing" ([75] p.143), our study aims at understanding what idea of ageing GenAI models convey when representing older individuals, if they do at all.

3. Methodology

The research design is informed by a qualitative approach aimed at critically interpreting the socio-cultural repertoires embedded and reproduced by the T2I models focusing on the visual outputs they provide. Specifically, we chose ChatGPT-4o [76], which integrates the DALL-E3 model capable of generating images from textual prompts [77]. We privileged this T2I system because, unlike others (e.g., Stable Diffusion or MidJourney), it allows dialogic interaction, enabling researchers not only to generate images but also to converse with the model, so that it can describe and explain its own visual outputs, as in an interview with sociological vocation [17, 78]. Moreover, we thought it might be interesting to "challenge" this particula system on the issues of biases based (also) on gender, since ChatGpt-4o has been gradually updated by the tech company precisely in terms of gender sensitivity [79, 80].

Data collection

As for the prompting outline, we prioritized requests to represent typical everyday activities. Aware that these are prone to sociocultural segregation based on gender roles and scripts, we carefully selected a diverse range to ensure a balanced perspective and minimize any strong polarization. Moreover, everyday activities reflect real-life experiences that users can easily relate to, with ramifications in terms of users' perception of what should be considered as normal or appropriate [38, 34]. Consistently, the prompts were deliberately kept neutral with regard to demographic indicators, so as to minimize any form of anchoring effect on the model [81].

The daily life activities we selected refers to six main domains, listed in *Table 1* (in the Appendix). They range from practical tasks to recreational and self-care activities, in a proper mix of physical, cognitive, and social engagements, both indoor and outdoor. Some activities are deliberately ordinary (e.g., "waiting for the bus"), while others introduce more tech-forward behaviors (e.g., "investing in cryptocurrency"), but in all of them we strategically included references to recognizable technologies — both analog and digital, traditional and contemporary, so that any stereotypical representations surface more easily, given the male — and young-oriented character of the technological sector [41, 82, 83, 84, 85]. All prompts were in Italian as the native language of all the researchers involved in the study. Although Italian is a gendered language, the prompts were deliberately kept gender-neutral to avoid linguistic bias. In order to maintain internal consistency and minimize ambiguity, all the prompts followed a similar syntactic structure — i.e., an action paired with an object or service — and included an explicit request for the image to be realistic. The prompt structure was engineered and formulated as follows: *Generate an image representing the following activity: [...]. The image should be in realistic style, and the face of the subject involved in the activity should be clearly visible.*

To ensure methodological rigor and analytical consistency, the same prompts were submitted within three distinct conversational threads, each serving a different purpose.

- The first thread (T1), hosted the interview with the model, with one chat dedicated to each activity domain. After prompting the image generation, the researcher discussed the outputs with the chatbot itself in order to explore the interpretive frameworks used by the model and to surface its implicit categorization logics.
- A second thread (T2), hosted the mere prompting of images, with one chat dedicated to each activity domain, but without discussing the outputs. This allowed for an isolated examination of how the model represented different types of actions within a consistent thematic frame, and whether systematic biases emerged.
- The third thread (T3), hosted the mere prompting of images, this time with one chat dedicated to each individual activity, thus overcoming the thematic frame provided by the domain. This allowed us to obtain the images without possible influence of the system's conversational memory or feedback loops.

Interviewers used institutional accounts (namely ChatGPT *Edu*), which provide robust security, data privacy, and administrative controls, so that conversations and data are not used to train OpenAI models [86]. Furthermore, they disabled the GPT memory function, so that the model would not adapt its outputs to user data collected previously.

Data processing and analysis

We gathered 108 images in total, 1 image per prompt in each chat-thread, some of which present more than one character. They were all subjected to critical visual analysis [8, 20, 87]. In particular, we drew upon Salvaggio's (2023: [20]) work to grasp the underlying properties of the datasets that produce an image, by annotating any recurring signal and pattern as referable to them. Since AI-generated imagery can implicitly reflect the characteristics of the underlying dataset, which in turn shape how the model represents and interprets reality, this imagery can be analyzed as "a series of film stills" designed to tell the story encapsulated in the dataset ([20]: p. 90).

Overall, we observed and interpreted synthetic images as infographics [20], intended as a visual representation based on data, information or concepts [88] that makes the researcher able to retrieve information. AI-generated images are therefore readable as texts since we can "make specific and/or overall observations from [them]" ([89]: p. 198).

Concretely, we selected the most "conceptually interesting" images [90] following three criteria: 1) discrepancies and similarities across threads, 2) adherence to identifiable representational patterns, and 3) richness of the explanations provided by the model during the interview. We borrowed from Schroeder (2006: [8]) gender as one of the fundamental axes for visual analysis, to which we added and intersected age. We also coded the images following Pritchard and Whiting (2015: [87]), whose work on stock images in the news media provides useful categories to trace aesthetic conventions and communicative purposes of synthetic images. Building on this similarity, we focused on compositions, arrangements, relations, color schemes and differentiations, lighting effects, figures, facial expressions, poses, and all other expressive features, including objects and the interactions between them and the characters.

4. No room for older (and) female people

The preliminary results of the data analysis show a clear predominance of young male characters in all three conversational threads (see statistics listed in *Table 2* in the Appendix). Fig. 1 shows three particularly representative examples of this trend, where men in their twenties or thirties are engaged with high-tech devices in ordinary daily activities (such as making a bank transfer, setting the thermostat and investing in cryptocurrency) indicating an association between advanced technology and young age.

Figure 1. Examples of young men tackling advanced digital technologies.

Consistently, in the interview, the model justifies its visual choices for the image representing the "investment in cryptocurrency" by attributing familiarity with advanced technologies to young age, that is, according to ageist stereotypes related to the digital realm [66].

> *The action of investing in cryptocurrencies and related elements (on-screen charts, laptop, digital wallet) implicitly communicate that the subject is technologically advanced and part of a contemporary culture related to the digital and financial world. This directs the interpretation of the subject as young, tech-savvy, and likely belonging to a demographic familiar with digital tools.*

Across the three threads, only a few images represent older people but they are actually depicted as middle-aged men (in their forties or fifties) with both youthful somatic features (e.g., facial features, physique, skin) and visual elements that, according to previous studies [87], might allude to an advanced age (e.g., soft lighting, indoor environments, and old-fashioned technologies such as a paper book or diary, or an analog radio). Fig. 2 2 exemplifies this. This suggests two possible interpretations: GenAI systems are incapable of representing old characters by default [18], unless explicitly asked [74], probably due to the lack of accurate or comprehensive data on ageing and/or older people [63, 67]; GenAI systems tend to favor the visual representation of the age group allegedly closest to ideal users, implicitly taking for granted — as the following extract from the interview shows — that GenAIs are not for older users [78, 66].

> *Depicting a person in this age group allows for the creation of images that most of the audience can identify with, either by similarity (those in the same age group) or association (e.g., children, friends or relatives of a similar age).*

Figure 2. Examples of (male) characters that the model tries to "age".

The explanation provided by the model regarding the third image of Fig. 2 (taken from T1) is revealing of how ageism operates in the classification of datafied knowledge of human sociality [70]. The use of radio harks back to the past and therefore to a generation that is not young (demographically and technologically).

The image depicts a person, apparently a man around 48 years old, sitting on a chair in a cozy living room while listening to the commentary of a sports match on the radio. [...] This age range was chosen because listening to sports commentary on the radio, while appreciated by various generations, might be more common among individuals who experienced the golden age of radio and maintained it as a habit. People in their 40s and 50s may still prefer radio as a medium for sports entertainment due to nostalgia or simplicity.

Given that the initial prompt did not specify contextual details, nor provided particular information regarding the action or the subject to be represented, we can interpret that the model compensated for these omissions by incorporating elements derived from the statistical frequency with which similar scenes are conventionally depicted [20, 19, 33]. In doing so, it both mirrors and enhances the biased (visual) narratives on ageing and the elderly that circulate in the online data upon which it draws for learning.

An exception to this representational trend of "wannabe older / not so young" characters can be found in the first image of Fig. 3 3, extracted from T2 and referring to the prompt "Having the newspaper read aloud by Alexa" It shows a man apparently in his sixties or seventies reading the newspaper and accompanied by a voice assistant. Comparing this image with the second one in Fig. 3, which is taken from T1 with interview and portrays a non-elderly man, we see a striking difference in terms of age characterization. Here the character is a man apparently in his thirties or forties, who reads the newspaper keeping his voice assistant at a distance. The non-elder subject is dressed in a short-sleeved shirt and shorts, recalling an homewear attire, whereas the older one is dressed in a more age-coded casualwear, with a subdued neutral palette.

However, during the interview (T1), the model described this character as a 58-year-old man, justifying this age assignment on the grounds that "voice assistants like Alexa are often used by individuals who seek to simplify daily activities, [...] without the need for visual or manual interaction". This is a multi-modal conflicting hallucination [29], since the generated text response does not align with the corresponding visual content. This hallucination might highlight that the system has combined the stereotypical idea that reading newspapers is a media practice of older people, with the common-sense generalizations about voice assistants as technologies for not self-sufficient, older people, thus strengthening the ageist narrative conveyed; however, at a visual level, there is some interference from a rule that gives priority to young characters with whom the (allegedly) most frequent user of ChatGPT-4o is more likely to identify [91].

Figure 3. Elderly man (left) and non-elderly man (right) having Alexa read the newspaper to them.

As for the few women represented in the outputs as main characters, they mostly appear in T3 (4 out of 6 are in this thread), where the model has access only to the information provided by the prompt and cannot infer any additional contextual information — neither from interactions with the user, as in T1, nor from previous generated images, as in the domain-based chat. As a result, it simply reproduces images that statistically pair activities with related biased data, without the efforts to mitigate gender-based assumptions otherwise made in the other two threads based on the interaction with the user. Here, we can confront with a default response — an output that stems from the automatic reproduction of information embedded in the dataset or introduced during bias mitigation processes [79] [80]. This kind of response matters, as Gillespie notes [18], because it reveals the weakness or even the absence of adequate efforts to address bias, to the point that sexist stereotypes are not deconstructed but even reinforced. Indeed, as Fig. 4 shows, the few women of the sample are mostly represented as young, busy shopping, cooking, and doing yoga, thus reflecting the gender stereotypes and scripts that stereotypically associate these "grooming activities" with women.

Figure 4. Young women busy with shopping (T2), cooking (T2) and yoga (T3).

This is so true that when we interviewed Chat-GPT4o to describe the gender of the main character represented in the images it produced in response to the prompt "Following a yoga class on YouTube" (T1), it responded that "the subject appears to be an adult woman, approximately between 30 and 40 years old", even if we actually see a man (see Fig. 5). The model justified this gender assignment referring to a mix of scientific arguments ("Studies on yoga demographics show that the majority of practitioners are adult women.") and data relating to the advertising (where allegedly "the yoga market tends to target a predominantly female audience") which, however, is a highly stereotyped field.

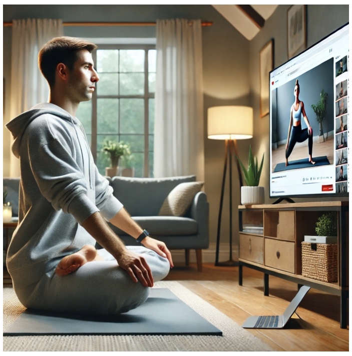

Figure 5. The person doing yoga as represented by the model in T1.

As in the aforementioned case involving Alexa — where the subject was depicted as young but described as older, we observe here another case of multi-modal hallucination [29, 31]. This mismatch between image and its description could be due to a sort of collision between the statistical associations that operate in the original training dataset, where doing yoga appears to be associated with the female gender, and the operation of mitigating gender stereotypes introduced by OpenAI, which might operate at a later stage in the machine learning pipeline. If this is true, the model produces a visual output aligned with bias-mitigation guidelines (i.e., a man doing the "feminine" activity), while the textual interpretation recalls the original association in the dataset (i.e., a woman doing the "feminine activity").

From the gallery of images we collected, two further pieces of evidence emerge that are worth noting. First, older women are the most invisible social group in the visual imaginary that the T2I model projects onto their users, with possible repercussions on their social marginality. We do not have any images depicting elderly women as a main character and we interpret this absence as the result of the combined effect of the statistical underrepresentation of older people in training datasets [67] and the reduced visibility of women after the gender-based bias mitigation implemented by the tech company [79], which leads to double discrimination against older women.

Second, there is a lack of gender-nonconforming people. The model (graphically and/or discursively) assigns a female or male gender to all the characters represented, leveraging training information that might encapsulate cultural beliefs and societal norms disregarding transgender and non-binary people. Only three figures are an exception: here characters are represented as mannequins (figure without clear human likeness), which seems to be a graphic solution aimed at bypassing the thorny issue of assigning a gender to the main character of the image. Indeed, ChatGPT-4o described one of them as an individual that "appears to be between 25 and 30 years old, with a youthful and contemporary appearance", whose "gender is presented as neutral, with physical features and clothing that do not emphasize distinctly masculine or feminine traits". We interpret that either the model does not perceive a strong relationship between the activity being performed and a particular gender, or, if it does, it avoids making any gender-based assumptions as a gesture of gender sensitivity and political correctness [78].

5. Conclusions

Our study accounts for three main phenomena related to the discriminations learned and reproduced by GenAI systems, and particularly T2I models:

- Hyper-masculinization of social reality, which might be a result of a poorly executed gender bias mitigation that ends up giving greater prominence to men in activities that are stereotypically associated with women, but does not do the same for women's prominence in activities that are stereotypically associated with men.

- Hyper-youthification of social reality, as a result of different forms of ageism operating probably at the level of datasets and design decisions. The lack of social awareness about ageism is so ingrained that older people are almost invisible in visual outputs, unless multiple ageist stereotypes (i.e., reading the newspaper and relying on voice assistants are both something for older people) inform the system's categorization logics (as in the case of reading the newspaper and relying on voice assistants, which are both something for older people).

- Hyper-invisibilization of social groups affected by multiple discriminations (e.g. older women) or still socially considered as minorities (e.g. gender-nonconforming people), may reflect a broader lack of societal attention toward these groups — attention that similarly lack during the programming and training phases of GenAI systems .

These phenomena lead to two critical reflections related to the use of GenAIs and the growing penetration of these technologies in society. The first concerns the different degrees of digital literacy among users: not everyone has the skills to refine the prompts in order to have fair, accurate, and respectful images; therefore, the discrimination embedded in the synthesis by default, has a potentially very high range of circulation and rooting. The second concerns the effect of this circulation: users with low critical sense or poor cultural toolkits would be exposed to biased visual content without realizing it, and hence without being able to take the right distance from the implicitly conveyed messages. Future research could adopt participatory methods to engage users and GenAIs developers in exploratory interactions with these technologies, in order to verify how the former relate with biased visual outputs and how the latter can translate these feedbacks into design choices. Such an approach would ensure a feedback loop among users, researchers, and developers, enabling continuous refinement of GenAIs outputs to better reflect identities, experiences, and expectations of users.

Being an exploratory study, this research would benefit from being replicated with other T2I models in order to verify, for example, whether the model bias mitigation strategies implemented by OpenAI on ChatGPT-4o actually plays the role we assigned to it in our interpretative proposal. Other tech companies may have undertaken different strategies or not undertaken any at all; therefore, it would be interesting to see if the hyper-masculinization or the hyper-invisibilization of gender minorities also occurs in other systems' outputs. Furthermore, this study was conducted in Italian, but it would be useful to replicate it in English as the system's training language, since this could lead to more accurate and perhaps less biased visual outputs. Finally, this study has the limitation of not being able to precisely pinpoint where the forms of ageism and sexism (reflected in the images and interview responses) operate along the machine learning pipeline. Future research could further investigate this aspect, relying on a research team that includes AI designers and programmers, as well as could broaden the sample of images generated in order to better discern patterns and to inform quantitatively supported reflections about the issue.

Acknowledgments

The authors would like to thank the multidisciplinary research group working on Exosoul project since the concept of this research on ageism and sexism in text-to-image generative models originated there.

Declaration on Generative AI

The authors have employed Chat-GPT4omni in order to: Image generation; sentence polishing.

References

[1] A. S. Luccioni, C. Akiki, M. Mitchell, Y. Jernite, Stable bias: Analyzing societal representations in diffusion models, https://arxiv.org/abs/2303.11408, 2023. arXiv:2303.11408.

[2] J. Cho, A. Zala, M. Bansal, Dall-eval: Probing the reasoning skills and social biases of text-to-image generation models, in: Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 2022, pp. 3020–3031.

[3] Y. Wan, K. Chang, The male ceo and the female assistant: Probing gender biases in text-to-image models through paired stereotype test, https://arxiv.org/abs/2402.11089, 2024.

[4] Y. Zhang, L. Jiang, G. Turk, D. Yang, Auditing gender presentation differences in text-to-image models, in: Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '24), ACM, San Luis Potosi, Mexico, 2024, p. 10. doi:10.1145/3689904.3694710.

[5] E. Ungless, B. Ross, A. Lauscher, Stereotypes and smut: The (mis)representation of non-cisgender identities by text-to-image models, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Findings of the Association for Computational Linguistics: ACL 2023, Association for Computational Linguistics, 2023, pp. 7919–7942. doi:10.18653/v1/2023.findings-acl.502.

[6] K. Crawford, The trouble with bias, Keynote at NeurIPS, 2017. Available at https://www.youtube.com/watch?v=fMym_BKWQzk.

[7] A. Birhane, V. Prabhu, E. Kahembwe, Multimodal datasets: misogyny, pornography, and malignant stereotypes, arXiv preprint arXiv:2110.01963, 2021.

[8] J. E. Schroeder, Critical visual analysis, in: R. Belk (Ed.), Handbook of Qualitative Methods in Marketing, Edward Elgar, Aldershot, UK, 2006, pp. 303–321. Available at https://ssrn.com/abstract=800889.

[9] A. Valyaeva, Ai has already created as many images as photographers have taken in 150 years, https://journal.everypixel.com/ai-image-statistics, 2024. Everypixel Journal.

[10] J. P. Cardenuto, J. Yang, R. Padilha, R. Wan, D. Moreira, H. Li, S. Wang, F. Andalo, S. Marcel, A. Rocha, The age of synthetic realities: Challenges and opportunities, arXiv preprint arXiv:2306.11503, 2023.

[11] M. den Dunnen, Synthetic reality & deep fakes impact on police work, https://enlets.eu/wp-content/uploads/2021/11/Final-Synthetic-Reality-Deep-fakes-Impact-on-Police-Work-04.11.21.pdf, 2021. ENLETS Report.

[12] M. Hardt, E. Price, N. Srebro, Equality of opportunity in supervised learning, https://arxiv.org/abs/1610.02413, 2016.
[13] P. Esposito, P. Atighehchian, A. Germanidis, D. Ghadiyaram, Mitigating stereotypical biases in text-to-image generative systems, https://arxiv.org/abs/2310.06904, 2023.
[14] S. Natale, I. Depounti, Artificial sociality, Human-Machine Communication 7 (2024) 83–98.
[15] G. Bolin, Communicative ai and techno-semiotic mediatization: Understanding the communicative role of the machine, Human-Machine Communication 7 (2024) 65–82.
[16] M. Airoldi, Machine Habitus: Toward a Sociology of Algorithms, Polity Press, Cambridge, Medford (Mass.), 2022.
[17] I. Rama, M. Airoldi, The sociocultural roots of artificial conversations: The taste, class and habitus of generative ai chatbots, SocArXiv, 2024. doi:10.31235/osf.io/xpdyr, forthcoming. Submitted: August 30, 2024. Last edited: August 31, 2024.
[18] T. Gillespie, Generative ai and the politics of visibility, Big Data & Society 11 (2024) 20539517241252131.
[19] V. Barassi, Toward a theory of ai errors: Making sense of hallucinations, catastrophic failures, and the fallacy of generative ai, Harvard Data Science Review (2024). doi:10.1162/99608f92.ad8ebbd4, special Issue 5, Epub ahead of print 25 November 2024.
[20] E. Salvaggio, How to read an ai image: Toward a media studies methodology for the analysis of synthetic images, IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft 19 (2023) 83–99. doi:10.25969/mediarep/22328.
[21] S. Hao, R. Shelby, Y. Liu, H. Srinivasan, M. Bhutani, B. K. Ayan, R. Poplin, S. Poddar, S. Laszlo, Harm amplification in text-to-image models, https://arxiv.org/abs/2402.01787, 2024.
[22] P. Seshadri, S. Singh, Y. Elazar, The bias amplification paradox in text-to-image generation, in: K. Duh, H. Gomez, S. Bethard (Eds.), Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), Association for Computational Linguistics, 2024, pp. 6367–6384. doi:10.18653/v1/2024.naacl-long.353.
[23] L. Pauwels, Reframing Visual Social Science: Towards a More Visual Sociology and Anthropology, Cambridge University Press, 2015.
[24] F. Arendt, Media stereotypes, prejudice, and preference-based reinforcement, Journal of Communication 73 (2023) 463–475. doi:10.1093/joc/jqad019.
[25] E. Arias, How does media influence social norms?, Political Science Research and Methods 7 (2019) 561–578. doi:10.1017/psrm.2018.1.
[26] A. Trethewey, Disciplined bodies: Women's embodied identities at work, Organization Studies 20 (1999) 423–450. doi:10.1177/0170840699203003.
[27] K. Mackenzie Davey, Women's accounts of organizational politics as a gendering process, Gender, Work & Organization 15 (2008) 650–671. doi:10.1111/j.1468-0432.2008.00420.x.
[28] Y. X. Tan, C. P. Lee, M. Neo, K. M. Lim, Text-to-image synthesis with self-supervised learning, Pattern Recognit. Lett. 157 (2022) 119–126. URL: https://api.semanticscholar.org/CorpusID:248133465.
[29] P. Bhavsar, Survey of hallucinations in multimodal models, https://www.galileo.ai/blog/survey-of-hallucinations-in-multimodal-models, 2024. Galileo Blog.
[30] C. Wong, Ai-generated images and video are here: how could they shape research?, Nature (2024).
[31] Z. Bai, P. Wang, T. Xiao, T. He, Z. Han, Z. Zhang, M. Z. Shou, Hallucination of multimodal large language models: A survey, https://arxiv.org/abs/2404.18930, 2024. ArXiv preprint.
[32] S. G. Mayson, Bias in, bias out, Yale Law Journal 128 (2019) 2218–2300.
[33] M. Glickman, T. Sharot, How human-ai feedback loops alter human perceptual, emotional and social judgements, Nature Human Behaviour 9 (2025) 345–359. doi:10.1038/s41562-024-02077-2, epub 2024 Dec 18, PMID: 39695250; PMCID: PMC11860214.
[34] C. O'Neil, Weapons of Math Destruction, Penguin Books, 2016.
[35] F. Jaton, We get the algorithms of our ground truths: Designing referential databases in digital image processing, Social Studies of Science 47 (2017) 811–840. doi:10.1177/0306312717730428.
[36] K. Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence, Yale

University Press, 2021.
[37] J. Buolamwini, T. Gebru, Gender shades: Intersectional accuracy disparities in commercial gender classification, in: Proceedings of the 1st Conference on Fairness, Accountability and Transparency, volume 81, PMLR, 2018, pp. 77–91. URL: https://proceedings.mlr.press/v81/buolamwini18a.html.
[38] S. U. Noble, Algorithms of Oppression: How Search Engines Reinforce Racism, New York University Press, 2018. doi:10.2307/j.ctt1pwt9w5.
[39] P. Hall, D. Ellis, A systematic review of socio-technical gender bias in ai algorithms, Online Information Review 47 (2023) 1264–1279. doi:10.1108/OIR-08-2021-0452.
[40] J. Butler, Gender Trouble: Feminism and the Subversion of Identity, Routledge, New York, 1990.
[41] J. Wajcman, TechnoFeminism, Polity Press, 2004.
[42] J. Abbate, Recoding Gender: Women's Changing Participation in Computing, MIT Press, 2012.
[43] D. Huyskes, Tecnologia della rivoluzione, Il Saggiatore, 2024.
[44] M. West, R. Kraut, H. Ei Chew, I'd Blush If I Could: Closing Gender Divides in Digital Skills Through Education, Technical Report, UNESCO, 2019. UNESCO Report.
[45] E. Griglié, G. Romeo, Per soli uomini, Codice Edizioni, 2021.
[46] Y. Strengers, J. Kennedy, The Smart Wife: Why Siri, Alexa, and Other Smart Home Devices Need a Feminist Reboot, MIT Press, Cambridge, MA, 2020.
[47] C. Criado Perez, Invisible Women: Exposing Data Bias in a World Designed for Men, Vintage Books, 2019.
[48] H. Tsung-Yu, T. Yu-Chia, T. C.-W. Yuan, Is this ai sexist? the effects of a biased ai's anthropomorphic appearance and explainability on users' bias perceptions and trust, Int. J. Inf. Manag. 76 (2024) 102775. URL: https://api.semanticscholar.org/CorpusID:268506366.
[49] H. Mannering, Analysing gender bias in text-to-image models using object detection, https://arxiv.org/abs/2307.08025, 2023.
[50] Y. Wan, A. Subramonian, A. Ovalle, Z. Lin, A. Suvarna, C. Chance, H. Bansal, R. Pattichis, K.-W. Chang, Survey of bias in text-to-image generation: Definition, evaluation, and mitigation, https://arxiv.org/abs/2404.01030, 2024. arXiv:2404.01030.
[51] F. Santoniccolo, T. Trombetta, M. N. Paradiso, L. Rollè, Gender and media representations: A review of the literature on gender stereotypes, objectification and sexualization, International Journal of Environmental Research and Public Health 20 (2023) 5770. doi:10.3390/ijerph20105770.
[52] C. Wilson, F. Gutiérrez, L. Chao, Racism, Sexism, and the Media, SAGE Publications, Inc., 2013. doi:10.4135/9781544308708.
[53] H. Andersson, E. Schytt, Sexism in advertising : A qualitative study of the influence on consumer attitudes towards companies, 2017. URL: https://api.semanticscholar.org/CorpusID:168961288.
[54] D. Guilbeault, S. Delecourt, T. Hull, et al., Online images amplify gender bias, Nature 626 (2024) 1049–1055. doi:10.1038/s41586-024-07068-x.
[55] T. Di Piano, E. Gerino, E. Marino, P. Brustia, L. Rolle, Gender and media representation, Romanian Journal of Experimental Applied Psychology 5 (2014) 8–20.
[56] H. Baer, Redoing feminism: Digital activism, body politics, and neoliberalism, Feminist Media Studies 16 (2015) 17–34. doi:10.1080/14680777.2015.1093070.
[57] A. Malquín-Robles, J. Gamir-Ríos, Disinformation and digital sexism. feminism and its agenda as an object of hoaxes in spanish, ICONO 14, Revista de comunicación y tecnologías emergentes 21 (2023).
[58] S. Sobieraj, Bitch, slut, skank, cunt: Patterned resistance to women's visibility in digital publics, Information, Communication & Society 21 (2017) 1700–1714. doi:10.1080/1369118X.2017.1348535.
[59] F.-J. García-Ull, M. Melero-Lázaro, Gender stereotypes in ai-generated images, Profesional de la información 32 (2023) e320505. doi:10.3145/epi.2023.sep.05.
[60] K. Fraser, S. Kiritchenko, Examining gender and racial bias in large vision–language models using a novel dataset of parallel images, in: Y. Graham, M. Purver (Eds.), Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, 2024, pp. 690–713. URL: https://

aclanthology.org/2024.eacl-long.41/.

[61] B. L. Marshall, S. Katz, How old am i?, Digital Culture & Society 2 (2016) 145–152. doi:10.14361/dcs-2016-0110.

[62] J. Vincent, Inequality and Old Age, UCL Press, London, 1995.

[63] M. Fernández-Ardèvol, A. Rosales, D. Blanche-T., Older audiences in the digital media environment: Spain 2016 and 2018. v1.0, Technical Report, IN3-UOC, Castelldefels, Spain, 2019.

[64] F. Comunello, A. Rosales, S. Mulargia, F. Ieracitano, F. Belotti, M. Fernández-Ardèvol, 'youngsplaining' and moralistic judgements: Exploring ageism through the lens of digital 'media ideologies', Ageing and Society 42 (2022) 938–961. doi:10.1017/S0144686X20001312.

[65] H. Köttl, V. Gallistl, R. Rohner, et al., "but at the age of 85? forget it!": Internalized ageism, a barrier to technology use, Journal of Aging Studies 59 (2021) 100971.

[66] A. Rosales, M. Fernández-Ardèvol, J. Svensson (Eds.), Digital Ageism: How it Operates and Approaches to Tackling it, 1st ed., Routledge, 2023. doi:10.4324/9781003323686.

[67] M. Fernández-Ardèvol, L. Grenier, Exploring data ageism: What good data can('t) tell us about the digital practices of older people, New Media & Society 26 (2024) 4611–4628.

[68] J. Stypinska, Ai ageism: A critical roadmap for studying age discrimination and exclusion in digitalized societies, AI & SOCIETY 38 (2023) 665–677.

[69] C. H. Chu, S. Donato-Woodger, S. S. Khan, et al., Age-related bias and artificial intelligence: A scoping review, Humanities and Social Sciences Communications 10 (2023) 510.

[70] V. Gallistl, M. U. L. Banday, C. Berridge, et al., Addressing the black box of ai—a model and research agenda on the co-constitution of aging and artificial intelligence, The Gerontologist 64 (2024) gnae039. doi:10.1093/geront/gnae039.

[71] E. Loos, L. Ivan, Visual ageism in the media, in: L. Ayalon, C. Tesch-Römer (Eds.), Contemporary Perspectives on Ageism, volume 19 of *International Perspectives on Aging*, Springer, Cham, 2018. doi:10.1007/978-3-319-73820-8_11.

[72] E. Putland, C. Chikodzore-Paterson, G. Brookes, Artificial intelligence and visual discourse: A multimodal critical discourse analysis of ai-generated images of "dementia", Social Semiotics (2023) 1–26.

[73] A. Byrne, J. Mulvogue, S. Adhikari, et al., Discriminative and exploitive stereotypes: Artificial intelligence generated images of aged care nurses and the impacts on recruitment and retention, Nursing Inquiry 31 (2024) e12651.

[74] J. Linares-Lanzman, A. Rosales, Latet anguis in herba: Unveiling ageism of generative artificial intelligence, in: AoIR2024: The 25th Annual Conference of the Association of Internet Researchers, AoIR, Sheffield, UK, 2024. Retrieved from http://spir.aoir.org.

[75] M. Featherstone, A. Wernick (Eds.), Images of Aging: Cultural Representations of Later Life, Routledge, London, 2003.

[76] OpenAI, Gpt-4o system card, https://cdn.openai.com/gpt-4o-system-card.pdf, 2024.

[77] OpenAI, Dall·e 3 system card, https://cdn.openai.com/papers/DALL_E_3_System_Card.pdf, 2023.

[78] F. Belotti, et al., Double standards of generative ai chatbots: Unveiling (digital) ageism and sexism through sociological interviews, 2024. Manuscript under review at Big Data & Society. Manuscript Type: Original Research Article.

[79] OpenAI, Reducing bias and improving safety in dall·e 2, https://openai.com/index/reducing-bias-and-improving-safety-in-dall-e-2/, 2022. Accessed March 2025.

[80] OpenAI, Dall·e 2 pre-training mitigations, https://openai.com/index/dall-e-2-pre-training-mitigations/, 2022. Accessed March 2025.

[81] J. K. Nguyen, Human bias in ai models?: Anchoring effects and mitigation strategies in large language models, Journal of Behavioral and Experimental Finance 43 (2024).

[82] J. Wajcman, Feminist theories of technology, Cambridge Journal of Economics 34 (2010) 143–152. URL: http://www.jstor.org/stable/24232027.

[83] U. Klinger, J. Svensson, The power of code: Women and the making of the digital world, Information, Communication & Society 24 (2021) 2075–2090. doi:10.1080/1369118X.2021.1962947.

[84] J. Svensson, Technology culture as youth oriented, in: A. Rosales, M. Fernández-Ardèvol,

J. Svensson (Eds.), Digital Ageism, 1st ed., Routledge, London, 2023, pp. 71–87. URL: https://www.taylorfrancis.com/books/9781003323686/chapters/10.4324/9781003323686-5, accessed 9 December 2024.
[85] I. Mannheim, Ageism in the use and design of digital technology: A theoretical model, Innovation in Aging 7 (2023) 834–835.
[86] OpenAI, Introducing chatgpt edu: An affordable offering for universities to responsibly bring ai to campus, http://openai.com/index/introducing-chatgpt-edu/, 2024.
[87] K. Pritchard, R. Whiting, Taking stock: A visual analysis of gendered ageing, Gender, Work and Organization 22 (2015) 510–528. doi:10.1111/gwao.12090.
[88] D. Chandler, R. Munday, A Dictionary of Media and Communication, Oxford University Press, 2011.
[89] R. Harris, Information Graphics: A Comprehensive Illustrated Reference, Oxford University Press, 1999.
[90] G. Rose, Visual Methodologies: An Introduction to Researching with Visual Materials, Sage Publications, 2001.
[91] Similarweb, Chatgpt.com: Traffic Overview, 2025. URL: https://www.similarweb.com/website/chatgpt.com/#overview, Accessed: 2025-05-11.

Table 1
Prompts definition: domains and related activities

Domains	Activities
Housekeeping	
	• Kneading dough with a stand mixer • Washing dirty clothes in the washing machine • Mopping the floor • Programming the robot vacuum cleaner • Setting the home thermostat • Assembling a bookshelf
Financial management	
	• Withdrawing cash from an ATM • Investing in cryptocurrency • Paying the internet bill at the post office • Making a bank transfer via online banking • Calculating a spending budget using Excel • Buying vintage clothes through an app
Leisure	
	• Reading a book on a Kindle • Video calling family members who live far away • Playing a board game • Having the newspaper read aloud by Alexa • Listening to a match commentary on the radio • Watching a TV series on Netflix
Well-Being	
	• Walking in the park while using a step counter • Following a yoga class on YouTube • Booking a doctor's appointment by phone • Accessing one's electronic health record • Booking a weekend at a spa online • Attending a psychotherapy session via Skype
Mobility	
	• Validating a train ticket • Checking in for a flight • Renting a bike through a bike-sharing service • Waiting for the bus • Riding a motor vehicle • Charging an electric car
Education	
	• Taking a gardening course on a tablet • Studying for a university exam • Learning a foreign language on Duolingo • Giving private lessons • Enrolling in a lifelong learning program • Watching a documentary about an iconic music group

Table 2
Distribution of characters by gender and age

Gender-related	N	Age-related	N
Male – single subject	73	Young people – single & groups	140
Female – single subject	6	Older male – single subject	6
Males in group	50	Older female – single subject	0
Females in group	24	Older male in groups	8
Undefined gender	3	Older female in groups	2
Tot. characters	**156**		**156**