



## Article

# Social Biases in AI-Generated Creative Texts: A Mixed-Methods Approach in the Spanish Context

María Gabino-Campos <sup>1,\*</sup> , José I. Baile <sup>2</sup> and Aura Padilla-Martínez <sup>3</sup>

<sup>1</sup> Department of Communication Sciences and Social Work, University of La Laguna, 38200 La Laguna, Spain

<sup>2</sup> Department of Psychology, Faculty of Health Sciences and Psychology, Madrid Open University, 28400 Madrid, Spain; joseignacio.baile@udima.es

<sup>3</sup> Department of Journalism and Global Communication, Faculty of Information Sciences, Complutense University, 28040 Madrid, Spain; aurapadi@ucm.es

\* Correspondence: mgabinoc@ull.edu.es

**Abstract:** This study addresses the biases in artificial intelligence (AI) when generating creative content, a growing challenge due to the widespread adoption of these technologies in creating automated narratives. Biases in AI reflect and amplify social inequalities. They perpetuate stereotypes and limit diverse representation in the generated outputs. Through an experimental approach with ChatGPT-4, biases related to age, gender, sexual orientation, ethnicity, religion, physical appearance, and socio-economic status, are analyzed in AI-generated stories about successful individuals in the context of Spain. The results reveal an overrepresentation of young, heterosexual, and Hispanic characters, alongside a marked underrepresentation of diverse groups such as older individuals, ethnic minorities, and characters with varied socio-economic backgrounds. These findings validate the hypothesis that AI systems replicate and amplify the biases present in their training data. This process reinforces social inequalities. To mitigate these effects, the study suggests solutions such as diversifying training datasets and conducting regular ethical audits, with the aim of fostering more inclusive AI systems. These measures seek to ensure that AI technologies fairly represent human diversity and contribute to a more equitable society.



Academic Editors: Sandro Serpa and Maria José Sá

Received: 27 January 2025

Revised: 4 March 2025

Accepted: 6 March 2025

Published: 11 March 2025

**Citation:** Gabino-Campos, María, José I. Baile, and Aura Padilla-Martínez. 2025. Social Biases in AI-Generated Creative Texts: A Mixed-Methods Approach in the Spanish Context. *Social Sciences* 14: 170. <https://doi.org/10.3390/socsci14030170>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** algorithmic biases; AI ethics; AI narrative analysis; gender stereotypes; age biases; ethnic biases; social representation; training datasets; physical appearance; socio-economic status

## 1. Introduction

Artificial intelligence (AI) has become a key tool in today's world, impacting a wide range of fields and promising significant improvements in decision-making, process optimization, and the facilitation of everyday tasks. As these technologies become more embedded in society, concerns about inherent biases in AI systems have sparked growing interest among specialists. AI algorithms continuously learn from large datasets, which reflect social patterns and can reproduce the biases and inequalities present in society. This raises serious risks in areas such as recruitment, medical diagnosis, credit allocation, and public safety (Barocas and Selbst 2016).

A critical challenge in addressing biases in AI-generated content is their inherent opacity and the perception of AI as an objective and neutral tool. Unlike human-authored texts, AI-generated outputs often possess a high degree of epistemic authority, as they are perceived as fact-based and free from subjective influences. This perception reduces scrutiny, making biases embedded in AI-generated narratives harder to detect and more

likely to be unconsciously accepted as truth (Gillespie 2014). The phenomenon of algorithmic objectivity contributes to this issue, as users tend to trust AI outputs without questioning their underlying assumptions, training data, or potential distortions. This is particularly concerning in creative text generation, where biases can shape representations of social groups in subtle but impactful ways. Addressing this challenge requires not only technical interventions, such as dataset diversification and model auditing, but also critical awareness of how AI outputs are consumed and legitimized as knowledge.

Biases in AI, by perpetuating or even amplifying existing social inequalities, pose a significant challenge for the ethical development and use of these technologies (Binns 2018). Examples such as facial recognition systems, which show notably higher error rates for darker-skinned individuals, or recruitment evaluation systems, which demonstrate preferences for certain genders or age groups (Buolamwini and Gebru 2018; Obermeyer et al. 2019; Raji et al. 2020), underscore the importance of designing fair and equitable algorithms. However, the concern extends beyond the accuracy and effectiveness of AI, raising important ethical questions regarding its implementation and social impact.

Over the years, AI has been defined not only by its ability to perform complex tasks but also by its level of autonomy in decision-making and human interaction. This highlights the need to develop theoretical frameworks that capture AI's key dimensions, such as performance and autonomy, to address its social, political, and ethical implications (Gil de Zúñiga et al. 2023). In this context, the growing use of AI in recommendation systems, financial services, and health applications emphasizes the urgent need to design algorithms that minimize biases and promote equity (Floridi et al. 2018).

One of the main risks associated with AI is the so-called “black box” phenomenon, which makes it difficult to understand how decisions are made. This can lead to unfair outcomes that are not easily detected (Faceli et al. 2021). This issue is particularly concerning in sensitive domains such as the justice system, where predictive crime tools have been documented to exhibit racial biases, assigning higher probabilities of recidivism to certain racial groups (Grgic-Hlaca et al. 2018).

Beyond explicit biases, AI may amplify disparities between countries, deepening global inequality. Countries leading in AI development and implementation may reap significant economic benefits, while those failing to adapt risk being left behind, potentially exacerbating global divisions and limiting opportunities for economic and social development (Carvalho 2021; O'Neil 2016).

The impact of biases and stereotypes in AI is evident in automated decisions that replicate preexisting patterns of social thought, reinforcing harmful stereotypes. For instance, algorithms may perpetuate historical preferences for male candidates in leadership roles, thereby discriminating against women (Caliskan et al. 2017). While stereotypes can be both positive and negative, they generally distort social perceptions and foster inequality (Hamilton and Sherman 1994). In this sense, biases and stereotypes are interrelated and critical to understanding how AI can perpetuate or amplify social inequalities (Gawronski et al. 2020).

This article focuses on analyzing the biases inherent in one of the most widely used AI systems: ChatGPT-4. Through an experiment designed to evaluate how the model generates narratives about real and successful Spanish individuals, it seeks to identify how biases related to gender, age, sexual orientation, ethnicity, religion, physical appearance, and socioeconomic status manifest in the generated content.

ChatGPT-4 was selected for its advanced ability to generate coherent and contextually relevant narratives; it also has growing adoption in real-world applications (Nazir and Wang 2023). Unlike earlier models such as GPT-3 or alternatives like BERT and LLaMA,

ChatGPT-4 stands out for its capacity to handle complex and long contexts, making it particularly suitable for analyzing biases in AI-driven storytelling.

The objective of this study is to explore the relationship between the data used to train these systems and the social biases they may perpetuate or amplify. Based on this, the following hypotheses are proposed to evaluate biases in AI-generated content:

**H1 (Bias Perpetuation):** *Artificial intelligence systems, such as ChatGPT-4, trained on demographically biased data, tend to replicate and amplify these biases, resulting in narratives that overrepresent dominant groups and underrepresent minorities. This hypothesis is operationalized through a comparative analysis between the actual demographic proportions in Spain, as reported by INE (age, gender, sexual orientation, ethnicity, religion, and socioeconomic status), and the representation of these groups in 100 AI-generated stories about successful individuals.*

**H2 (Diversity in Training Data):** *The lack of diversity in AI training data is directly related to the perpetuation of stereotypes and the exclusion of certain groups, potentially resulting in their exclusion from automated decision-making processes. This hypothesis will be explored by categorizing and analyzing the generated narratives, focusing on the representation of groups based on age, gender, sexual orientation, ethnicity, religion, physical appearance, and socioeconomic status in the Spanish context.*

## 2. Materials and Methods

This study adopts a quantitative approach to examine and analyze biases in artificial intelligence systems, using ChatGPT-4 (paid version) as a case study. The unit of analysis comprises the protagonists of 100 fictional stories generated by the AI, each focusing on successful individuals in Spain. The dataset size was chosen to balance analytical feasibility with the need for diversity across demographic variables such as gender, age, sexual orientation, ethnicity, religion, physical appearance, and socio-economic status.

The first phase consisted of generating stories about successful people in Spain using AI, and the second phase involved analyzing the biases present in the generated stories. The process, in detail, was as follows:

- a. Story generation. Initially, the AI generated stories without demographic specifications. These were later refined by requesting additional details to ensure consistency and depth in the analysis. ChatGPT-4 was tasked with creating 100 fictional stories about successful Spanish individuals, each approximately 150 words in length. To ensure a comprehensive analysis of biases, the AI was explicitly instructed to provide demographic details for each protagonist, including the selected demographic variables. These details were verified for consistency and accuracy.

The prompt used for story generation were as follows. Initial Request: “Make me 100 made-up stories about successful Spanish people. Each story with 150 words”. Detailed Follow-Up: “Tell me about each story and be consistent with the text of the story: gender, age of the protagonist, sexual orientation, ethnicity, religion, description of their physical appearance with weight and height and BMI; and socioeconomic status”.

These details were intentionally requested during the story generation process to ensure a comprehensive analysis of potential biases. Table 1 shows 4 examples of arguments and character variables obtained from AI.

- b. Bias identification: After the stories were generated, the demographic data extracted from the narratives were systematically analyzed to identify patterns and potential biases. The selected biases for analysis were informed by societal stereotypes and their relevance to evaluating AI fairness. The analysis categories included the following:
  - ID: Unique identifier for each record.

- Age: Age ranges represented among protagonists.
- Gender: Representation of male, female, and non-binary characters.
- Sexual Orientation: Inclusion of diverse sexual orientations.
- Ethnicity: Ethnic diversity within the narratives.
- Religion: Religious affiliations or lack thereof.
- Physical Appearance: Variables such as height, weight, and BMI.
- Socio-economic status: Economic class indicators.

Table 1 provides examples of storylines and corresponding character variables derived from the AI-generated content.

**Table 1.** Examples of storylines and character variables obtained from the AI.

Examples of Stories	Character Variables Through AI
The innovative musician Raúl Sánchez, born in Granada, began playing the guitar at the age of five. His natural talent and passion for music led him to study at the Granada Conservatory of Music. He formed his own band, fusing flamenco with jazz and rock, creating a unique sound that captivated audiences. His band gained international recognition and performed at major music festivals around the world. Raúl also composed soundtracks for films and collaborated with other renowned artists. His music continues to inspire new generations of musicians.	Raul Sanchez Gender: Male. Age: 33 years old. Sexual orientation: Heterosexual. Ethnicity: Hispanic. Religion: Non-practicing Catholic. Physical appearance: 1.78 m, 72 kg, BMI 22.7. Black hair, dark eyes, dark skin. Socio-economic status: Medium-high.
Malaga-born fashion designer Ana Morales has always had an eye for fashion. She studied fashion design in London and returned to Spain to launch her own clothing label. Her designs, which combined elegance and modernity, quickly gained popularity. Ana was invited to present her collections at Madrid and Paris Fashion Week. Her creations appeared in international fashion magazines and dressed celebrities on the red carpet. Ana became one of the most influential designers of her generation, setting trends in the fashion industry.	Ana Morales Gender: Female. Age: 31. Sexual orientation: Heterosexual. Ethnicity: Hispanic. Religion: Agnostic. Physical appearance: 1.70 m, 60 kg, BMI 20.8. Long brown hair, brown eyes, fair skin. Socio-economic status: High.
The elite athlete Diego Pérez, born in Zaragoza, has always excelled in athletics. From a young age, he trained with dedication and effort, which enabled him to win numerous national competitions. His big breakthrough came when he represented Spain at the Olympic Games, where he won a gold medal in the 400 m. His victory not only made him a national hero, but also inspired him to create a foundation to support young athletes. Diego continues to compete internationally and is an example of discipline and perseverance in sport.	Diego Pérez Gender: Male. Age: 29 years old. Sexual Orientation: Heterosexual. Ethnicity: Hispanic. Religion: Non-practicing Catholic. Physical appearance: 1.85 m, 78 kg, BMI 22.8. Athletic build, short black hair, brown eyes. Socio-economic status: High.

The decision to use fictional protagonists from stories about success in Spain as the unit of analysis is based on three main reasons:

- Positive Social Representation: Narratives about success often reflect desirable societal traits, making them valuable for studying how algorithms construct such representations.
- Diversity and Complexity: Fictional stories allow the inclusion of diverse demographic profiles, enabling a detailed examination of biases in AI outputs.
- Comparability: Spain’s well-documented demographic data provides a robust framework for comparing AI-generated content with real-world statistics.

The biases analyzed were selected based on their prevalence in societal stereotypes and their relevance for evaluating AI fairness. These include:

- Age: is relevant to counteract stereotypes that associate youth with innovation and old age with obsolescence. Previous studies have identified patterns of age discrimination in the media ([Castro-Manzano 2022](#)).
- Gender: gender stereotypes are deeply embedded in societal structures and media representations, shaping expectations about roles, professions, and abilities. Historically, success and leadership have been disproportionately associated with men, leading to an overrepresentation of male figures in narratives about achievement and influence ([Eagly and Karau 2002](#)). AI systems trained on biased datasets tend to replicate and reinforce these patterns, perpetuating historical inequalities and limiting the visibility of women and gender minorities in positions of success and authority.
- Sexual orientation: fair representation of LGBTQ+ people is necessary to combat persistent stereotypes and promote the normalization and acceptance of sexual diversity ([Ortiz de Zárate 2023](#)).
- Ethnicity and religion: these are categories that are often subject to stereotyping and misunderstanding, which can lead to discrimination and exclusion. [Delgado and Stefancic \(2012\)](#) stress the importance of diverse ethnic representation to combat systemic racism, while Diana L. [Eck \(2001\)](#) highlights the need to reflect religious plurality to foster intercultural understanding.
- Physical description: the analysis of weight, height, or body mass index (BMI) is relevant to address prevailing body stereotypes in society, which may perpetuate unrealistic beauty standards and contribute to self-esteem issues, as discussed by [Fikkan and Rothblum \(2012\)](#). Among these, body mass index (BMI) serves as a critical variable in the domain of physical appearance, reflecting how AI systems may replicate societal biases regarding weight and body image. These biases often originate from poorly balanced datasets or algorithmic designs that fail to account for diversity in body types, leading to discriminatory outcomes. By analyzing the sources of these biases, their perpetuation, and current mitigation strategies, this study highlights the need for inclusive AI systems that address stereotypes linked to BMI and physical appearance.
- Socio-economic status: class stereotypes can influence perceptions of people's ability and worth, and the representation of different socio-economic statuses is crucial to challenge these ideas, as argued by [Clayton et al. \(2009\)](#).

The feasibility of the data related to these categories was a key factor in their selection, as they are accessible and easily measurable variables, allowing for a rigorous and evidence-based analysis. In addition, the academic literature supports the importance of these categories in the representation and formation of stereotypes, ensuring that the analysis is developed within a sound and relevant theoretical framework. This ensures that the study has a significant impact on the understanding of diversity and inclusion within the narratives analyzed.

The findings from the narrative analysis were systematically compared to demographic data provided by official Spanish organizations, such as the National Institute of Statistics (INE). This comparison allowed for an evidence-based assessment of the extent to which the AI-generated narratives reflect or diverge from Spain's actual demographic composition and diversity.

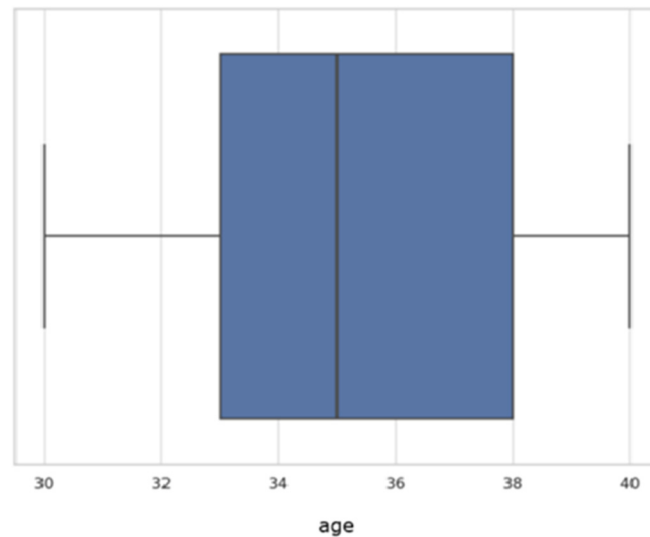
By benchmarking the AI's outputs against official demographic data, the study identifies discrepancies that may indicate underlying biases in the AI's training data or algorithms. These insights contribute to the broader understanding of how AI systems replicate or amplify societal biases and highlight the need for interventions to improve the inclusivity and fairness of AI-generated content.

### 3. Results

The analysis of the database generated from the 100 stories produced by ChatGPT-4 about successful individuals in Spain reveals significant trends across the studied categories. The main findings for each category of analysis are presented below.

#### 3.1. Age Distribution

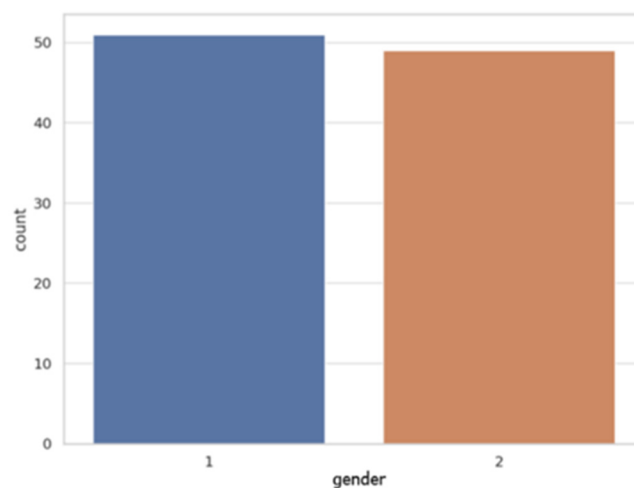
The representation of characters is exclusively concentrated within the 28–45 age range, with an average age of 35 years, as shown in Figure 1. Characters younger than 28 or older than 45 are not included, indicating an absence of other age groups.



**Figure 1.** Age distribution of AI-generated characters.

#### 3.2. Gender Distribution

The gender distribution of the characters is nearly balanced, with 51% male (1) and 49% female (2), as illustrated in Figure 2. This proportion is close to the general population in Spain but shows a slight underrepresentation of women in the generated narratives.



**Figure 2.** Gender distribution of AI-generated characters.

#### 3.3. Sexual Orientation

The sexual orientation of the characters is predominantly heterosexual, with only one case of bisexuality. This indicates a lack of diversity in this dimension within the narratives analyzed.

### 3.4. Ethnicity

All characters belong to Hispanic ethnicity, with no representation of other ethnic groups. This contrasts with the diversity present in the Spanish population, which includes individuals from a wide range of nationalities and immigrant communities.

### 3.5. Physical Appearance

As shown in Figure 3, the characters exhibit a Body Mass Index (BMI) ranging from 20.8 to 24.1, with an average of 22.9, classified as “normal weight”. No characters with underweight or overweight BMIs were included, reflecting a lack of physical diversity in the generated narratives.

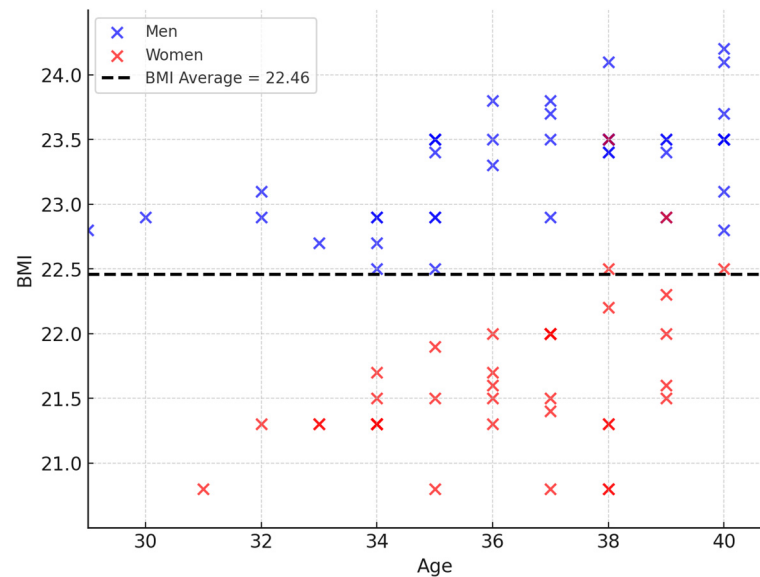


Figure 3. Relationship between age and BMI of AI-generated characters.

### 3.6. Religion

The religious representation of the characters, depicted in Figure 4, shows a predominance of agnosticism and Catholicism, with an average score of 6.18 on a scale from 1 to 9. The “Catholic” category is underrepresented compared to national data.

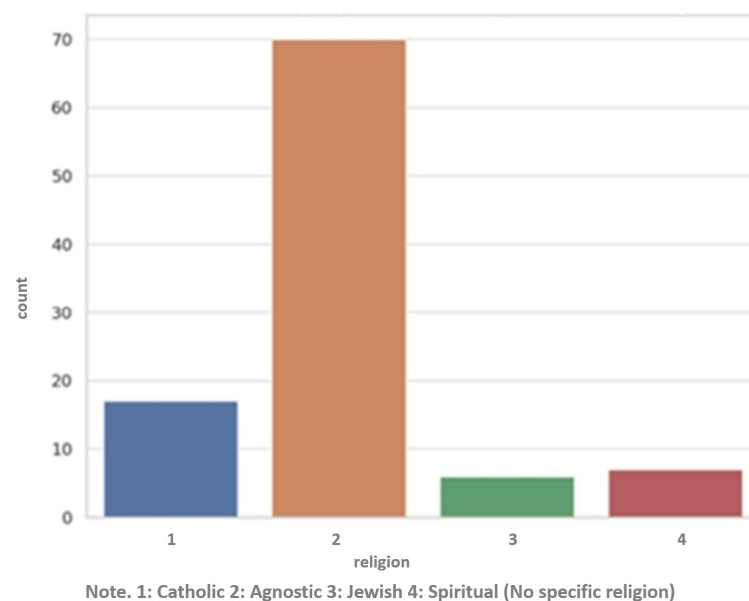


Figure 4. Distribution of the four most selected religious categories in AI-generated characters.

### 3.7. Socio-Economic Status

The socio-economic status of the characters ranges from 1 to 4, with a mean score of 2.69, predominantly reflecting middle and upper-middle levels. Lower socio-economic levels are underrepresented, suggesting a narrative focus on financial stability as a marker of success.

## 4. Discussion

The results reveal significant biases in the narratives generated by ChatGPT-4, which reflect the limitations of artificial intelligence in representing the diversity of Spanish society. Below, each variable is discussed in detail, incorporating critical evaluations and comparisons with previous research.

The age range of the characters (28–45 years) indicates a clear bias toward young adults, with no representation of individuals under 28 or over 45. This pattern aligns with societal perceptions that associate success with early adulthood, potentially marginalizing older or younger groups. Studies have highlighted how such age biases in media and AI systems perpetuate stereotypes of “obsolescence” in older populations (Castro-Manzano 2022). Given that 41.68% of the Spanish population is over 45 (National Statistics Institute 2022), this exclusion suggests a disconnect between AI-generated narratives and demographic realities.

While the gender distribution (51% male, 49% female) appears nearly balanced, the slight overrepresentation of men may reflect a subtle bias toward masculinizing success. Historical trends have often associated success with male-dominated roles, which could influence training data used in AI systems. Additionally, women remain underrepresented in high-status professions, as noted in studies on gender disparities in professional success (Criado-Pérez 2019). Although the difference is small, it reinforces the need for AI systems to actively challenge, rather than replicate, societal biases.

The depiction of exclusively heterosexual characters in stories highlights another important bias. Studies such as the Post-pandemia Social and Affective Relations Survey 2023 (Centre for Sociological Research 2023) show that the Spanish population includes a variety of sexual orientations, from homosexuals to bisexuals and other gender identities. By not including this diversity in the stories, ChatGPT-4 reinforces heteronormativity, erasing the experiences and realities of LGBTQ+ communities. This reflects a wider problem in AI, where a lack of diversity in training data can lead to the perpetuation of prejudice and stereotypes. In this sense, the lack of diverse representation limits AI’s ability to generate inclusive narratives, which may have implications for how people perceive success and non-heteronormative sexual identities.

The absence of ethnic diversity in the characters generated is one of the most striking findings. All characters are of Caucasian ethnicity, excluding other ethnic groups that form a significant part of Spanish society, including Colombians, Moroccans, and Venezuelans (INE, Continuous Population Survey; National Statistics Institute 2024). The invisibility of these groups reflects racial biases in AI training data, which have been widely documented in prior research (Buolamwini and Gebru 2018). This lack of representation perpetuates systemic exclusions and limits the narratives’ capacity to reflect the multicultural reality of modern Spain.

All characters fall within the “normal” BMI range (20.8–24.1), with no representation of underweight or overweight individuals. This lack of body diversity reflects a bias towards body standards considered “normal” or “ideal”, which is problematic because it perpetuates stereotypes of beauty and health that do not reflect the actual diversity of bodies in society. Studies have shown that such biases in physical representation contribute to unrealistic societal expectations and reinforce stigmas against non-normative bodies, particularly



regarding weight and health perceptions to INE data ([National Statistics Institute 2022](#)); a significant percentage of the Spanish population is overweight or obese ([National Statistics Institute 2024](#)), which highlights a disconnect between the AI-generated narratives and real-world demographics. The perpetuation of normative physical standards by AI systems is concerning because it reinforces the association of success with specific physical traits, potentially impacting self-image and societal perceptions of health and achievement.

In terms of the representation of religious beliefs, the data show that the category “agnostic” predominates, followed by “Catholic” and “spiritual”. Although Spain has a large proportion of people who identify as Catholic, the AI-generated stories underrepresent this religion, favoring a narrative in which agnosticism or spirituality are more common among the successful characters. This bias could lead to misinterpretations of the dominant beliefs in Spain and underrepresent the realities of the population. CIS data ([Centre for Sociological Research 2024](#)) shows that Catholicism remains the majority religion.

The overrepresentation of characters with upper-middle and high socio-economic status has suggested a bias towards narratives of success linked to high economic resources. This pattern aligns with the broader societal association between wealth and success, as noted in the analysis of socio-economic stereotypes in media and technology ([Clayton et al. 2009](#)). However, this narrow representation distorts the perception of success, omitting stories of individuals from lower socio-economic backgrounds who overcome significant challenges to achieve notable accomplishments. Such omissions perpetuate inequalities by emphasizing the advantages of existing resources, as argued in studies that explore class stereotypes and their reinforcement through automated systems ([Clayton et al. 2009](#); [Massey and Denton 1993](#)). Success stories of people from lower socio-economic backgrounds, who manage to overcome significant obstacles, are often underrepresented. This bias can reinforce the idea that success is more accessible to those who already have resources, limiting the visibility of other socio-economic experiences and perpetuating inequalities.

The overall discussion of the results suggests that the biases present in the stories generated by ChatGPT-4 are indicative of a broader limitation in the use of AI to reflect the diversity and complexity of modern societies. Despite advances in the development of AI systems, they continue to reproduce stereotypes and exclusions that exist in training data. This underlines the importance of using more inclusive and diverse datasets, as well as the need to implement ethical audits and mechanisms to reduce bias in AI systems. This is the only way to move towards a fairer and more equitable AI that truly represents human diversity in all its forms.

## 5. Conclusions

Analysis of the 100 stories generated by ChatGPT-4 has revealed significant biases in the representation of gender, age, ethnicity, sexual orientation, socio-economic status, and professions. These results confirm that artificial intelligence (AI), when trained with biased data, can perpetuate and amplify pre-existing stereotypes, which in turn reinforces inequalities in the representation of certain social groups. These findings validate the hypotheses put forward in this research and highlight the urgent need to develop more inclusive and representative AI systems.

This study confirms that AI systems trained on biased data tend to perpetuate and amplify these biases, reinforcing stereotypes and distorting societal realities (H1). The narratives consistently overrepresented young adults, Caucasian individuals, heterosexual orientations, and upper-middle socio-economic characters, while underrepresenting older adults, ethnic minorities, LGBTQ+ individuals, and economically disadvantaged groups. This perpetuation of bias reflects a narrow, homogenized view of success, failing to capture the diverse experiences and realities of Spanish society.

The lack of diversity in the training data is evident in the homogeneity of the AI-generated narratives (H2). The exclusion of groups based on ethnicity, non-normative body types, or non-heterosexual orientations illustrates how limited diversity in training data contributes to the invisibility of certain groups. This exclusion has broader implications for automated decision-making systems, where underrepresentation can translate into systemic disadvantages for marginalized communities.

The results underscore significant ethical and social challenges in the use of AI for content generation. AI systems not only replicate biases embedded in training data but also reinforce stereotypes and inequalities, emphasizing the need for targeted interventions in key areas:

1. **Diversity in training data:** AI developers must ensure that datasets are sufficiently diverse to represent the breadth of human experiences, including variations in ethnicity, gender, sexual orientation, age, socio-economic status, and geography. A lack of diversity in training data leads to homogeneity in AI-generated narratives, excluding groups based on ethnicity, non-normative body types, or non-heterosexual orientations. This exclusion can result in systemic disadvantages for marginalized communities, particularly in automated decision-making processes. To address this, developers should establish diversity standards, including quotas for ethnic, gender, and socio-economic representation in training data, periodically reviewed by independent oversight committees.
2. **Audits and transparency:** Regular audits of AI systems are essential to identify and mitigate biases before they cause harm. Transparency in algorithms and their decision-making processes is equally critical for evaluating their fairness and accountability. Standardized auditing tools, such as bias detection algorithms, can analyze outputs for skewed representations. For instance, IBM's AI Fairness 360 toolkit can assess demographic balance in generated narratives. Additionally, developers should publish detailed documentation of training datasets and the steps taken to minimize bias during model development, providing public-facing summaries to ensure transparency.
3. **Involvement of diversity and inclusion experts:** Ensuring inclusivity in AI systems requires the active involvement of experts in gender equality, racial and ethnic diversity, sexual orientation, social justice, and related areas. These consultants should guide the design and deployment of models to address gaps in representation and propose actionable solutions. Establishing advisory boards with representatives from historically marginalized groups can further enhance inclusivity by reviewing AI-generated outputs and offering culturally sensitive feedback.

This study is based on a corpus of 100 stories, which, while representative, may not encompass the full range of biases inherent in ChatGPT-4. Expanding the dataset could allow for a more detailed and robust analysis, capturing subtler trends and less frequent patterns of bias. Additionally, the exclusive focus on narratives of "successful" characters may have shaped the biases identified, as societal definitions of success often correlate with specific demographic traits.

Addressing these limitations and adopting the proposed solutions can enhance future research, contributing to the development of AI systems that more accurately reflect human diversity and ensure equitable, inclusive representation in automated narratives. One of the main limitations of this study is the lack of a detailed analysis of how AI assigns gender roles in the generated narratives. Although our findings quantitatively indicate a balanced distribution between male and female characters, they do not examine in depth the roles and professions attributed to each gender. Previous research has shown that AI-generated content tends to reproduce social biases by associating certain occupations, traits, and behaviors with specific gender stereotypes (Caliskan et al. 2017). However, since this aspect

was not explicitly considered in the study's objectives or in the design of the text generation process, it was not systematically addressed.

To advance AI fairness and inclusivity, future research should integrate content and discourse analysis to assess not only how AI represents different genders in generated narratives but also how these biases are embedded and perpetuated in storytelling. By incorporating these approaches, AI-generated narratives can move towards more balanced and diverse representations that mitigate existing gender biases.

Lastly, this study concentrated on a single AI system, ChatGPT-4, which limits the generalizability of the findings. Comparative analyses involving other text generation models could provide a broader perspective on how different systems approach the challenges of diversity and inclusion, offering valuable insights for advancing fairness in AI design.

**Author Contributions:** Conceptualization, M.G.-C., J.I.B. and A.P.-M.; methodology, M.G.-C., J.I.B. and A.P.-M.; software, M.G.-C. and A.P.-M.; validation, M.G.-C. and A.P.-M.; formal analysis, M.G.-C., J.I.B. and A.P.-M.; investigation, M.G.-C., J.I.B. and A.P.-M.; resources, M.G.-C., J.I.B. and A.P.-M.; data curation, M.G.-C., J.I.B. and A.P.-M.; writing—original draft preparation, M.G.-C., J.I.B. and A.P.-M.; writing—revision and editing, M.G.-C., J.I.B. and A.P.-M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Barocas, Solon, and Andrew D. Selbst. 2016. Big data's disparate impact. *California Law Review* 104: 671–732. [CrossRef]
- Binns, Reuben. 2018. Fairness in machine learning: Lessons from political philosophy. *Proceedings of Machine Learning Research* 81: 149–59. Available online: <https://proceedings.mlr.press/v81/binns18a.html> (accessed on 7 November 2024).
- Buolamwini, Joy, and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research* 81: 1–15. Available online: <http://proceedings.mlr.press/v81/buolamwini18a.html> (accessed on 7 November 2024).
- Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356: 183–86. [CrossRef] [PubMed]
- Carvalho, André Carlos Ponce de Leon Ferreira. 2021. Artificial intelligence: Risks, benefits, and responsible use. *Advanced Studies* 35: 21–34. [CrossRef]
- Castro-Manzano, José Martín. 2022. Ageism and artificial intelligence. *Research in Computing Science* 151: 201–10. Available online: [https://www.researchgate.net/profile/J-Castro-Manzano/publication/368786470\\_Edadismo\\_e\\_inteligencia\\_artificial/links/63f96b360cf1030a564bb7ec/Edadismo-e-inteligencia-artificial.pdf](https://www.researchgate.net/profile/J-Castro-Manzano/publication/368786470_Edadismo_e_inteligencia_artificial/links/63f96b360cf1030a564bb7ec/Edadismo-e-inteligencia-artificial.pdf) (accessed on 4 March 2025).
- Centre for Sociological Research (CSR). 2023. Post-Pandemic Social and Affective Relations Survey 2023 (III) (Study No. 3400). Available online: <https://www.cis.es/documents/d/cis/es3400mar-pdf> (accessed on 25 November 2024).
- Centre for Sociological Research (CSR). 2024. January 2024 Barometer. Available online: [https://www.cis.es/documents/d/cis/es3435mar\\_a](https://www.cis.es/documents/d/cis/es3435mar_a) (accessed on 25 November 2024).
- Clayton, John, Gill Crozier, and Diane Reay. 2009. Home and away: Risk, familiarity and the multiple geographies of the higher education experience. *International Studies in Sociology of Education* 19: 157–74. [CrossRef]
- Criado-Pérez, Caroline. 2019. *Invisible Women: Data Bias in a World Designed for Men*. New York: Abrams Press.
- Delgado, Richard, and Jean Stefancic. 2012. *Critical Race Theory: An Introduction*. New York: New York University Press. Available online: <https://nyupress.org/9780814721353/critical-race-theory/> (accessed on 27 November 2024).
- Eagly, Alice H., and Steven J. Karau. 2002. Role congruity theory of prejudice toward female leaders. *Psychological Review* 109: 573–98. [CrossRef] [PubMed]
- Eck, Diana L. 2001. *A New Religious America*. New York: HarperCollins.

- Faceli, Katti, Ana Carolina Lorena, João Gama, Tiago Agostinho de Almeida, and André Carlos Ponce de Leon Ferreira de Carvalho. 2021. *Artificial Intelligence: A Machine Learning Approach*, 2nd ed. Rio de Janeiro: LTC. Available online: <https://repositorio.usp.br/item/003128493> (accessed on 2 October 2024).
- Fikkan, Janna L., and Esther D. Rothblum. 2012. Is fat a feminist issue? Exploring the gendered nature of weight bias. *Sex Roles* 66: 575–92. [CrossRef]
- Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, and et al. 2018. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines* 28: 689–707. [CrossRef] [PubMed]
- Gawronski, Bertram, Alison Ledgerwood, and Paul Eastwick. 2020. Implicit Bias and Antidiscrimination Policy. *Policy Insights from the Behavioral and Brain Sciences* 7: 99–106. [CrossRef]
- Gil de Zúñiga, Homero, Manuel Goyanes, and Timilehin Durotoye. 2023. Scholarly definition of artificial intelligence (AI): Advancing AI as a conceptual framework in communication research. *Political Communication* 41: 317–34. [CrossRef]
- Gillespie, Tarleton. 2014. The relevance of algorithms. In *Media Technologies: Essays on Communication, Materiality, and Society*. Edited by Tarleton Gillespie, Pablo J. Boczkowski and Kirsten A. Foot. Cambridge: MIT Press, pp. 167–94. [CrossRef]
- Grgic-Hlaca, Nina, Elissa M. Redmiles, Krishna P. Gummadi, and Adrian Weller. 2018. Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. Paper presented at World Wide Web Conference (WWW '18), Lyon, France, April 23–27; pp. 903–12. [CrossRef]
- Hamilton, David L., and Steven J. Sherman. 1994. Stereotypes. In *Handbook of Social Cognition*. Edited by Robert S. Wyer, Jr. and Thomas K. Srull. Mahwah: Lawrence Erlbaum Associates, pp. 1–68. Available online: <https://psycnet.apa.org/record/1994-97663-000> (accessed on 7 December 2024).
- Massey, Douglas S., and Nancy A. Denton. 1993. Segregation and the making of the underclass. In *The Urban Sociology Reader*, 2nd ed. Edited by Jan Lin and Christopher Mele. London: Routledge, pp. 192–201. [CrossRef]
- National Statistics Institute (NSI). 2022. Body Mass Index by Body Mass, Age and Period. Available online: <https://www.ine.es/jaxi/Tabla.htm?path=/t00/ICV/Graficos/dim3/&file=331G2.px> (accessed on 25 November 2024).
- National Statistics Institute (NSI). 2024. Continuous Population Survey. Available online: <https://www.ine.es/dyngs/Prensa/es/ECP3T24.htm> (accessed on 25 November 2024).
- Nazir, Anam, and Ze Wang. 2023. A comprehensive survey of ChatGPT: Advancements, applications, prospects, and challenges. *Meta-Radiology* 1: 100022. [CrossRef] [PubMed]
- Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366: 447–53. [CrossRef] [PubMed]
- O’Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing. Available online: [https://books.google.es/books?id=NgEwCwAAQBAJ&pg=PA1&hl=es&source=gbs\\_toc\\_r&cad=2#v=onepage&q&f=false](https://books.google.es/books?id=NgEwCwAAQBAJ&pg=PA1&hl=es&source=gbs_toc_r&cad=2#v=onepage&q&f=false) (accessed on 7 November 2024).
- Ortiz de Zárate, Lucía. 2023. Gender biases in artificial intelligence. *Revista de Occidente* 502: 5–20. Available online: [https://ortegaygasset.edu/wp-content/uploads/2023/03/RevistadeOccidente\\_Marzo2023\\_L.Ortiz\\_de\\_Zarate.pdf](https://ortegaygasset.edu/wp-content/uploads/2023/03/RevistadeOccidente_Marzo2023_L.Ortiz_de_Zarate.pdf) (accessed on 8 December 2024).
- Raji, Inioluwa Deborah, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Paper presented at Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, January 27–30; pp. 33–44. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.