



Exploring stereotypes, bias, and expectations of women in the open data context

Lorena R. Romero-Domínguez, Andrea Sixto-Costoya & Antonia Ferrer-Sapena

To cite this article: Lorena R. Romero-Domínguez, Andrea Sixto-Costoya & Antonia Ferrer-Sapena (12 Jun 2024): Exploring stereotypes, bias, and expectations of women in the open data context, *Gender, Technology and Development*, DOI: [10.1080/09718524.2024.2354103](https://doi.org/10.1080/09718524.2024.2354103)

To link to this article: <https://doi.org/10.1080/09718524.2024.2354103>



Published online: 12 Jun 2024.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



RESEARCH ARTICLE



Exploring stereotypes, bias, and expectations of women in the open data context

Lorena R. Romero-Domínguez^a , Andrea Sixto-Costoya^{b,c,d} and Antonia Ferrer-Sapena^c

^aDepartamento de Periodismo II, Facultad de Comunicación, Universidad de Sevilla, Grupo Democ-Media, Sevilla, Spain; ^bDepartamento de Trabajo Social y Servicios Sociales, Universitat de València, Grupo UISYS, Valencia, Spain; ^cUC3M-UAM3, Unidad asociada al Instituto Interuniversitario de Investigación Avanzada sobre Evaluación de la Ciencia y la Universidad (INAECU), Madrid, Spain; ^dInstituto Universitario de Matemática Pura y Aplicada – IUMPA, Universitat Politècnica de València, Valencia, Spain

ABSTRACT

This study aims to provide a qualitative analysis of the knowledge, beliefs, and perspectives of prominent women working in different sectors in the context of open data. The analysis is based on semi-structured interviews with an outline developed for the interview script. The content obtained was analyzed by attending to categories that deal directly with a gender perspective, covering existing biases, and possible modes of transformation through actions at individual, collective, and institutional levels that incorporate gender sensitivity when working with open data. The findings show that gender bias persists in the process of collecting, analyzing, and interpreting data, as well as in the formation of work groups. The solution, highlighted by all interviewees, involves the development of interoperable and consistent data collection and analysis models that integrate a gender perspective right from the outset, as well as the training of professionals in fields associated with the use of data. Central to this process is the role of government agencies, which should promote efficient public policies giving greater visibility to the roles of women in areas of knowledge associated with open data. Finally, this research also points to the need for innovations such as artificial intelligence to overcome these challenges.

ARTICLE HISTORY

Received 10 August 2023
Accepted 6 May 2024

KEYWORDS

Open data; gender perspective; bias; artificial intelligence

Introduction

Open data and the gender perspective are two concepts that have followed parallel trajectories in recent years. As Ruggieri et al. (2021) suggests, gender equality and open access are two priorities on the public agenda (Comisión Europea, 2020; UNESCO, 2014) that have often been analyzed independently despite the fact that they share a number of common goals, such as inclusivity, transparency, and sustainability. One consequence of this is that open access initiatives have failed to take into account the

CONTACT Andrea Sixto-Costoya andrea.sixto@uv.es Departamento de Trabajo Social y Servicios Sociales, Universitat de València, Grupo UISYS, Valencia, Spain.

importance of including either the sex variable or the gender variable in a systematic and differentiated way in activities involving data collection, analysis, or accessibility.

Despite the rapidly increasing prominence of open data practices in recent years, especially since the paradigm shift brought about by the rise of the Open Science movement and the establishment of a new dimension of open access in scientific research (Sociedad Max Planck, 2003), public administration (Ferrer-Sapena et al., 2011), and the business sector (Helbig et al., 2021; Herala et al., 2019; OpenCorporate, 2023), the integration of a gender perspective into data catalogs and the collection of disaggregated data has not developed at the same rate, reflecting a lack of progress in the implementation of improvements in this sector (Open Data Watch, 2021, 2022).

It is important to note that this problem is not limited to open data but is a structural issue with multiple causes that has been identified in various sectors. A clear example can be found in the field of science, where it has been repeatedly observed that the failure to disaggregate data by sex or gender skews results and, in the case of studies of health issues, undermines diagnostic accuracy (Álvarez-Díaz, 2020; Gogos et al., 2019; S. K. Lee, 2018; Spector & Overholser, 2019). To tackle the negative effects of this bias, numerous institutions (including the European Commission, the World Health Organization, the National Institutes of Health, and the United Nations) have stressed the importance of including gender as a distinctive variable in research (Barr & Temkin, 2022; European Commission, 2021, 2022; United Nations 2022b; World Health Organization, 2018).

In the realm of business innovation, especially in cases where data are used as raw material, it has also been demonstrated that gender bias can lead to poor practices that affect everything from design to subsequent product development. An example of this can be observed in the development of artificial intelligence (AI), and more specifically, machine learning, as biased data collection will produce skewed or incomplete models (Castillo, 2023). Wang et al. (2019), for example, show how an algorithm learned to associate images of shopping and kitchens with women. Similarly, Otterbacher et al. (2017) found that internet search engines were more likely to retrieve images of women if the search included terms such as “sensitive” or “emotional,” while searches including terms traditionally associated with “hard skills” yielded results with more images of men. Another example can be found in a study by Bolukbasi et al. (2016), who trained an AI tool using Google News to solve the analogy “man is to computer programmer what woman is to x.” The automatic response generated by the tool was “x=homemaker.”

Manasi et al. (2022) also point out the specific cases of voice assistants and robotics. In the case of the first, they note that most of these products are represented with feminine features in submissive roles (e.g., Cortana, Alexa, or Siri). In the case of the second, the authors observe how some jobs traditionally occupied by women are being gradually robotized, such as supermarket cashiers or receptionists.

Cirillo et al. (2020) suggest that technologies involving AI constitute a “double-edged sword” in that they can perpetuate gender and sex inequalities but also have the potential to mitigate them through the management of biases. They point to the case of developing precise AI systems in healthcare designed to avoid undesirable discriminatory tendencies. In this regard, it is necessary to differentiate desirable biases from undesirable biases. While desirable bias refers to purposefully considering gender and/

or sex differences in order to obtain an accurate diagnosis for both basic and clinical research, undesirable bias occurs when claims regarding sex or gender and medical conditions are made despite a lack of comprehensive evidence to support them or based on biased evidence. The most common problem is the lack of a representative sample of the population in the training data set, which may be due to social or cultural factors, often affecting groups that are historically discriminated against, such as women or ethnic minorities. Similarly, Lee et al. (2022) highlight the importance of taking into account sex and gender differences in the field of dermatology in machine learning algorithms, arguing that there are important differences in both the epidemiology and clinical presentation of dermatological conditions and stressing the need to design machine learning tools that ensure gender and sex equity by distinguishing between desirable and undesirable biases.

In addition to the absence of sex/gender variables, various studies and reports have also made it clear that there is still a lack of gender balance on work teams, and that concepts such as the “glass ceiling” and the “sticky floor” continue to be relevant today, both within and beyond the world of science (Black, 2016; Cho et al., 2014; Domínguez Prats & Cuesta, 2019; Erren et al., 2014; European Commission, 2021; Eurostat, 2019; Prages, 2009; Sixto-Costoya et al., 2022; United Nations, 2022a).

In recent years, efforts to bring open data and a gender perspective together have led to initiatives that aim to address the challenge of uniting these two concepts in order to offer solutions aimed at reducing the gaps and biases identified above. Examples of this include the work carried out by the United Nations Statistics Division, which has published a document providing the methodological and analytical information necessary to improve the availability, quality, and use of statistics with a gender perspective (United Nations, 2016), and the World Bank’s “Gender Data Portal,” which offers indicators for a range of topics including health, education, violence, and employment, allowing users to download data in various formats, and also to access them via an API (The World Bank, 2023).

Despite such initiatives, Open Data Watch has reported that sex/gender disaggregated data still represent only a very small proportion of all data available. The figures offered by this organization suggest that there is plenty of room for improvement: Of the 187 countries included in the report, 21 do not yet offer disaggregated statistics (Open Data Watch, 2021). In its report titled *Transforming the Data Landscape: Solutions to Close Gender Data Gaps*, it also indicates that only half of the possible gender indicators of the Sustainable Development Goals (SDGs) are disaggregated by sex (Open Data Watch, 2022, p. 13).

The various reasons behind the continued existence of the gender gap in this context (Grantham, 2020) include the following factors: insufficient financial and technical resources and trained personnel; a lack of political commitment (despite investments made by many governments); the underuse of disaggregated data, and a lack of training on how to use them. Another factor is the lack of consistent guidelines for data disaggregation, which hinders comparisons between countries due to their different data collection standards (Tandon & Pareek, 2019). All these factors help explain why inequalities persist in areas such as education, economic opportunities, healthcare, political participation, and security, with public policies that are ineffective for the achievement of full equality between women and men.

In view of the above, and in the interests of contributing to the definition of the problem and the identification of possible solutions, the aim of this study is to offer a qualitative analysis of the knowledge, beliefs, and perspectives on the future of prominent women working in various fields in the context of open data from a gender perspective, addressing the existing gap(s) in the open data sector. No previous study of this kind has been found in the literature. To develop a better understanding of the opportunities offered by the integration of a gender perspective into open data, which in turn will facilitate the creation of effective policies for the eradication of inequality, this study aims to answer the following research questions:

RQ1: How are gender inequalities experienced in professional environments associated with open data?

RQ2: What are the best practices for asserting the role of women and promoting their visibility in areas of knowledge associated with open data?

RQ3: What strategic actions (both individual and collective) can be adopted to obtain gender-sensitive data using big data and AI?

Methodological design

Interviews probably constitute the most widely used qualitative research method in the social sciences (Creswell, 2012; Nathan et al., 2019; Olson, 2011). This study assesses their usefulness for identifying the perspectives, interpretations, perceptions, and feelings of respondents when reflecting on their relationship with the open data sector (either as students or as professionals) and how their status as women has affected that relationship. The objective is thus to go beyond mere quantification or classification in order to explore their personal experiences and opinions (Corbetta, 2007, p. 344; Curasi, 2001, p. 1; Dimond et al., 2012, p. 1; Kelly, 2010, p. 309).

This research uses qualitative interviews as a way of making sense of the knowledge drawn from the meanings the respondents (re)construct during the interaction (Edwards & Holland, 2020, p. 582). The narrative that emerges from their answers provides the foundation for a corpus that can help to identify gender stereotypes and biases, while also affirming the role of women in the open data sector and promoting their visibility.

Of the various interview models used in scholarly research (Díaz-Bravo et al., 2013), the semi-structured interview was chosen for this study since it offers an intermediate degree of flexibility between fully structured interviews, with predetermined questions and specific answers from respondents, and open interviews, which are flexible and adaptable to each interviewee, establishing a dialogue that is “spontaneous, focused, and of variable intensity” (Gaínza Veloso & Canales Cerón, 2006, pp. 219–220), but where there is a risk of digressing from the research objectives (Meho, 2006, p. 1284).

A basic outline was developed for the interview script according to the research questions (academic and professional background, views on gender biases, and expectations/strategic actions in the data sector), but other questions were personalized or customized based on the academic background and field of practice of the respondent in each case. Similarly, the wording, order, and number of questions were adapted

to each interviewee, covering the same general topics but without a single predetermined sequential structure (Tonon, 2012, p. 55).

To minimize interviewer biases and assumptions that could influence both the choice of questions and the subsequent interpretation of the data (Edwards & Holland, 2020; Nathan et al., 2019), the interviews were assessed by two different experts, one from the field of open data management and the other from gender studies. Their assessments were based on criteria of sufficiency, clarity, coherence, and relevance, adapted from the model proposed by Escobar-Pérez and Cuervo-Martínez (2008), with a rating of 1 to 4, where: 1=Does not meet the criterion; 2=Low level; 3=Moderate level; 4=High level. Questions that scored 2 or lower were deleted. Others with moderate scores were modified and/or regrouped according to the recommendations provided by the two assessors.

The interviews were conducted in online video conferences using the Zoom application. Online interviews have now become a widely consolidated practice, especially since the global COVID-19 pandemic (Caetano & Moran, 2021; Gruber et al., 2020), as they resolve the time, location, and mobility issues associated with face-to-face meetings (Janghorban et al., 2014, p. 1). Compared to other methods of synchronous communication (chats, instant messaging, forums, etc.), videoconferencing more closely replicates the characteristics of a face-to-face interview because it enables the researcher to collect both verbal and non-verbal information from the interviewee on the screen (Irani, 2019, p. 3). In addition, it provides a context in which both the interviewer and the interviewee can add their interpretations of the observations made and allows a more spontaneous flow of information (O'Connor & Madge, 2001, p. 10.1).

Purposive sampling was used to select the interview subjects (Nathan et al., 2019, p. 401) in the interests of obtaining testimonies from women working in a range of professional fields directly related to open data use/management/production/coding/dissemination. This includes fields such as engineering, design, data science, academic research, data journalism, public management and administration, and activism.

The respondents were contacted via an email that provided an outline of the objectives of the research and a form to obtain their consent to the recording of audiovisual content and the assignment of copyright/image rights/interpretation rights, as confirmation that their participation was strictly voluntary (Lancharro Taverro et al., 2018, p. 2).

The profiles of the women interviewed are described in Table 1.

The interviews took place between the months of April and August 2022. Once transcribed (without the use of automated transcription), each interview was read through several times. The transcription was deposited in the Zenodo repository (Romero-Domínguez et al., 2024). An initial dynamic was then established, which involved grouping units of meaning into general categories in order to identify the topics that emerged from the answers provided. This approach is based on a recommendation by Lopezosa et al. (2022, p. 10), who argue that it is "important for the code to be broad enough to cover all answers given, but at the same time concise enough to be specific and useful."

The system of categories is shown in Table 2. Three general categories were defined in accordance with the sequential structure of the interviews (Training and Experience; Gender Biases; Expectations/strategic actions) and in line with the research questions. Within each of these a number of subcategories were identified to describe the main

Table 1. Interviewee profiles.

Code	Sector	Experience
E01	Political organizations and advocacy groups	Establishment of initiatives both in the field of open data, and in the promotion of the visibility of women in technology; participation in public policy development.
E02	Private sector and academic sector	Data visualization consulting for various companies and universities.
E03	Private sector, nonprofit sector, and partnering with universities	Research management at an NGO formed by several prestigious institutions; data coach in various clinics.
E04	Nonprofit sector and partnering with universities	Founding of a technological consulting organization with a gender perspective, providing a legal and public policy perspective.
E05	Publishing sector and partnering with universities	Economic journalism and journalistic research, working for publications with recognized impact, founder of a media organization that uses and promotes data journalism.
E06	Teaching, research, and science policy	Research in the area of open data and open science; participation on several of the sector's leading committees and working groups.
E07	Public administration	Coordination at a government agency of various projects associated with technology; responsible for pioneering initiatives in the area of open data portals.

Table 2. Category tree.

Category 1 (C1) <i>Gender perspective</i> (RQ1 and RQ2)	C1.1 Training and experience	Concepts: inequality, obstacles, difficulties
	C1.2 Gender biases	Concepts: definition, type, identification, awareness, personal experience, perpetuation of stereotypes, identification of responsibilities
	C1.3 Modes of transformation	Concept: gender mainstreaming
Category 2 (C2) <i>Expectations</i> (RQ3)	C2.1 Future actions	Concepts: visibilization, individual or collective action, institutional support
	C2.2 Opportunities	Concepts: challenges, definition of policy, assessment of level of implementation, AI actions

subjects discussed by the interviewees. Key concepts were also considered in approximate terms to streamline the coding process.

C1.1 includes indicators related to their academic background, experience, and development, and how being a woman has shaped them as professionals in the open data sector (RQ1). C1.2 deals directly with the gender perspective, covering existing biases, and possible modes of transformation through actions incorporating gender sensitivity when working with open data in an AI context (RQ2). Finally, C2 refers to expectations and strategic actions for the future, with references to ways of addressing current challenges in this area on individual, collective, and institutional levels (RQ3).

With the aim of ensuring intercoder reliability, the analysis protocol was tested using the pilot test proposed by Neuendorf (2002). An initial draft was applied informally through the independent coding of two interviews. This pre-testing made it possible to rewrite the coding instructions to clarify the measurement of the variables.

Results

As a first overview of the results, Figure 1 presents a summary of the coding results, indicating the interviewees whose responses addressed each of the concepts defined in the categories (highlighted in green). It is worth noting that only one respondent provided a definition of her understanding of gender bias, although all respondents

		Interviewee							
		E01	E02	E03	E04	E05	E06	E07	
Category 1 <i>Gender Perspective</i>	C.1.1 Training and Experience	Concepts							
		Inequality							
		Obstacles							
		Difficulties							
	C.1.2 Gender biases	Concepts							
		Definition							
		Type							
		Identification							
		Awareness							
		Personal experience							
		Perpetuation of stereotypes							
		Identification of responsibilities							
	C.1.3 Modes of transformation	Concept							
Gender mainstreaming									
Category 2 <i>Expectations</i>	C.2.1 Future actions	Concepts							
		Visibilization							
		Individual action							
		Collective action							
		Institutional support							
C.2.2 Opportunities	Concepts								
	Challenges								
	Definition of policy								

Figure 1. Table summarizing responses by category/interview.

recognized that they had been victims of such bias in their academic and professional lives (RQ1). When talking about best practices for asserting the role of women in areas of knowledge associated with open data (RQ2), there is a unanimous response in relation to the importance of visibilization initiatives as an essential means of fighting against gender bias. There was also a broad consensus on the inequality between men and women in relation to the identification of responsibilities, and on the challenges that still need to be tackled. In regard to RQ3, most of the interviewees recognize that there is still work to be done to face the challenges generated by Big Data and Artificial Intelligence in the open data sector. As stated by interviewees AI is not the panacea for avoiding the replication of biases, especially due to the fact that algorithms are “trained” with data that offer an inaccurate reflection of reality and exacerbate the inequality between men and women.

Also, the word cloud (Figure 2) shows us that the term “Data” was the most frequently used, with 534 repetitions. “Women” (169), “Gender” (109) and “People” (105) were also frequently used, but to a lesser extent. This reflects the significant impact that data technology has on society and the importance of examining how the open data sector has been shaped and is evolving, particularly in relation to a gender perspective. Other significant terms include “Project(s)” and “Research,” due to the urgent need of encouraging data collection that considers variables related to sex and gender. This facilitates the subsequent processing of data taking into consideration the specificities linked to these variables. Finally, the term “Journalism” can also be highlighted, as it plays a crucial role as an intermediary between open data and citizens. It appears in the middle of the list and is repeated 50 times.

As for the content of the interviews, it was categorized into the thematic sections presented below under separate headings, which are organized according to the research questions. The interviewees’ answers are cited in both direct and indirect quotes, accompanied by the interviewee code as indicated in Table 1.



Figure 2. Word cloud of most frequently used words (30 times or more).

C1 gender perspective

C1.1 training and experience

All of the interviewees have completed post-secondary studies, and a high percentage (nearly 85%) have master's degrees or PhDs. It is also notable that Sciences (E03, E07) and Engineering (E01) are not the only fields that appear to connect these women to the open data sector, as they also include professionals with training in the fields of Social Sciences and Law (E02, E04, E05) and Humanities (E06). Characteristics shared by all interviewees include involvement in field-specific work teams in their current professional activity, as well as international study and work experience. Another aspect that most respondents share, despite their different professional backgrounds, is their recognition of open data as clearly beneficial to the public, for use as an economic initiative (E07, E01), for obtaining and making sense of information (E02, E05), or for participating in political decision-making (E04), among other benefits. In this respect, all the interviewees made reference to social responsibility in open data in one context or another, to its association with a culture of transparency (E01, E05) and to its role in enhancing the quality of democracy (E03).

When discussing the topic of their professional experience, all of the interviewees remarked that the decisions they have adopted over the course of their careers have resulted in their working with data in one way or another. This evolution in their careers has been the product of various factors, including their identification at some point of the importance of the use of data in different areas, ranging from the academic sector (E06) to the private (E01, E04, E05) and government sectors (E07).

Significantly, several respondents remarked that when they started working with data, everything, from concepts to infrastructure, was still in the development process, and they did not realize that their work was in fact contributing to the definition of a new professional field. This ambiguity in relation to open data as a work area is common to all fields in which the respondents have professional experience, highlighting their role as pioneers in this sense. Of particular interest on this point are the contributions of respondents E02 and E05, who commented that the work they were doing with data was even viewed somewhat strangely in their respective professional fields

(E05: “I was a bit crazy about Excel in the newsroom, and I remember colleagues asking me: ‘What are those little cells?’”). One respondent also highlighted her curiosity and how this trait could be channeled into data analysis (E07: “I’ve always liked to ask myself why things happen. I mean, it’s something that... I’m a very curious mind”).

C1.2 gender biases

When interviewees were asked about their perception of how being a woman has shaped them as professionals in the open data sector, the answers were all very similar: All mentioned the need to keep working to achieve full equality between men and women in this area. This view was founded on the conviction that significant obstacles still exist for women with an academic and/or professional interest in data science. Respondents made indirect references to the “glass ceiling,” the “sticky floor,” or the “imposter syndrome” when discussing the difficulties that they have had to overcome throughout their professional lives and during their time as students. They identified various situations as examples of these inequalities, including the low numbers of women in certain fields of study, especially STEM (E01: “Currently, 13% of graduates in programs like computer science or telecommunications are women”; E04: “Trainers [...] are the main ones who reproduce the big problem we have in trying to get more female technologists interested in ITCs”), the lack of visibility of female role models in formal and informal educational contexts (E02: “in practical terms, and if you look at the conferences, the percentage of women is still much, much lower”), the low proportion of women in the digital world (E04: “in startups there are more or less 70% men and 30% women. But we don’t know how many women are developing technology”), and the lack of digitalization of female work environments (E01, E04).

Also identifiable in their responses was a criticism of the lack of women in managerial and decision-making positions (e.g., in the design of technological and algorithmic infrastructures, as E01 and E04 emphasize), especially when they were asked about the application of regulations being insufficient to eliminate the problems described above (E05: “Is it meaningful to say that the European Commission’s 40% target has been reached? Or should we be saying whether women are really gaining power at the biggest companies?”).

There was consensus among interviewees that a major consequence of these imbalances has been the establishment of male-dominated work environments in the open data sector and that this would therefore be a first step that would need to be addressed. This situation affects the extent to which women are willing and able to continue working in the sector (E01, E04). As they point out, it is not that women do not understand the technology, but that they need strategies to be able to deal with the lack of other women in these environments. In such cases, competition takes precedence (E02: “I believe that here too there is a factor of competitiveness that is of no use at all”) over other types of actions related more to cooperation and the “soft skills” associated with women (E01, E04). These skills enable women to create communities where they are encouraged to participate and where they support each other (E02, E07, E05). These include groups such as R Ladies, PyLadies, #Wethewomenjournalistsstop (associated with the International Women’s Day demonstrations in Spain), and the *Chicas Poderosas* network in Latin America.

These male-dominated environments replicate the gender biases present in offline environments. All interviewees acknowledge that the open data sector is vulnerable to

social stereotypes present in other contexts, and this has a direct impact on widening the digital gender gap. There was also a consensus that technology is designed by and for men (E01, E04), with serious deficiencies for women seeking to use it for their professional development. The interviewees pointed out that the fact that women were less likely to use these tools reduced their job opportunities in a male-dominated digital environment:

E01: Obviously, it's not the technology, it's the social aspect. The social roles of men and women have been transferred to the digital environment, but the burden it places on us as professionals is brutal. Because most women aren't using technologies so much to develop their careers, for key professional issues.

C1.3 modes of transformation

Most of the interviewees took the view that it is not data in themselves that have a gender perspective, but the analysis, research, and consultation carried out with those data.

E06: Data have no gender. Data are the key to being able to conduct research with a gender perspective and research with a gender perspective is driven by a policy of responsible research (...).

Although data have no gender, they need to be disaggregated to be able to incorporate a sensitivity to the different effects that a particular issue can have on men and women. The respondents recognized that the impacts are very different and that they need to be studied separately for each of these populations (E01, E03) in order to gain a more accurate understanding of the reality (E04). In this sense, the interviewees saw it as essential to develop mechanisms that can facilitate the identification of variables that take sex into account when data are collected, preparing the data from the beginning so that a gender perspective can be applied later in the modeling and analysis stages (E03, E04, E07). What is needed, then, is the incorporation of gender mainstreaming into every stage of the data life cycle, from collection through to publication. Only in this way will it be possible to design social policies that can contribute to real equality.

Despite this consensus, there was also a clear recognition among the respondents of the deficiencies of data catalogs in relation to disaggregation by sex, with data referring to women being largely invisible (E03: "the problem here is that not all databases are divided by sex from the outset"; E04: "There is a lack of focus from the beginning because the data are often collected by men who do not understand that X is not the same as Y..."). As E07 points out, "According to the UN, in 2020 only 39% of the data required to monitor the implementation of the SDGs took into account the gender perspective, which is necessary." As one of the interviewees acknowledges (E01), the problem lies in the existence of fields of knowledge traditionally associated with masculinized environments that have received more funding and that therefore have more advanced automation and datafication processes. On the other hand, data related to gender equality have been traditionally overlooked in the development of digitalization. This issue brings the question of the third digital divide back to the center of the debate, this time with a focus on the quantity of sex-disaggregated data

catalogs. If this situation is not corrected, the respondents suggest that “it will be more difficult to explore policies on the equality of women” (E01).

A key example offered by many of the respondents when discussing this problem is violence against women. Interviewees point to the need to make more information available on this issue, because the lack of existing information undermines the development of an accurate picture of the reality (E03), and in the most extreme cases may even result in a denial of its existence (E05). It is therefore important to allocate funding for the collection of data in this area (E01) and for government initiatives to give visibility to the projects being implemented (E07). Two of these projects were highlighted by some of the respondents: “Data against the Noise” (developed in Spain by Gender Data Lab), a project that uses the archives of the Spanish court system to identify the presence of misogynistic violence without resorting to statistical language and with the support of visual data (E04); and “*Unidas*,” a project developed by Data Pop Alliance in partnership with the German Development Agency to measure patterns of domestic gender-based violence during lockdowns in Mexico, Colombia, and Brazil.

As mentioned above, the interviewees argue that governments have an important role and responsibility as policymakers, which in the case of research, for example, involves establishing the gender perspective as a priority for project funding (E06: “I think it’s great that you have to fill that out. Because when you come across the question, you think about the answer. If you don’t come across the question you just keep going and think that it’s not relevant”). In the case of public administration, interviewees talk about taking gender into account right from the data collection stage and “even before, in the definition of public policy” (E07). This means that the public sector needs to provide more disaggregated data, and the subsequent incorporation of a gender perspective is essential to guide public policy aimed at combating inequalities (E01, E07). All interviewees recognize the efforts made to integrate gender mainstreaming at all stages of the data life cycle, from data collection to publication.

C2 expectations

C2.1 future actions and C2.2 opportunities

In relation to the future, all of the interviewees suggested that to be effective the changes need to reach all actors involved at all levels. There was also agreement on the vital role of public policy, with institutional action that guarantees the visibility of women and their work in the area of open data, promoting projects and funding plans (especially with the [EU] Next Generation Program, as E01 emphasizes) that meet the needs of women in the context of global programs. E07 adds in this respect: “the data obtained as a result of the projects contracted by the government agency should be made openly available, as these data offered openly should include this necessary gender perspective.”

In the academic sector, better data management infrastructures (E04) and clear and simple guidelines for research staff are needed, as well as incentives and encouragements (E06). Similarly, in relation to training and solutions that can benefit all parties, respondents also identified a need for awareness-raising (E07), education (E04), and

incentives not only in universities (faculty, research staff, students), but also for public employees and all professionals engaged in data analysis (E04, E07), so that they can understand the data holistically (E03) and ensure that they reflect reality. On this last point, respondents talked about the need for diversity on technical teams, as well as the inclusion on work teams of members of the communities that the data refers to, especially women, when issues of equality are being addressed. Interviewees E03 and E04 share two significant experiences in this area:

E03: If I want to make an app about menstruation, why—and I'm giving you this example because it's a real example—why is the team that created this app made up only of men? I mean, they have no idea what's going on.

E04: The example I wanted to give actually is a person developing an app about dating. (...) Well, if this app had been or was designed while taking into account female users, they would probably have extra functions that would allow other female users to be aware of, "hey, this user such-and-such," or "hey, that user so-and-so," and this is a shortcoming that the technology has.

Another key aspect to consider when thinking about a gender perspective in open data and future actions is the potential of artificial intelligence (AI) in relation to the elimination of biases in data. In this regard, even while acknowledging the limitations that the artificial intelligence/open data binary still suffers from at this stage, respondents pointed to its potential as a generator of opportunities for society and the importance of working out how to take advantage of it to reduce inequalities of all kinds, and especially those related to gender (E03: "I think we can do very interesting things from the outset with machine learning technology and with the limited data that may exist in some regions").

Among the precautions to consider, interviewees mentioned the technical limitations (algorithms that can be improved, as E01 points out), the need for more training and for the establishment of rights and ethics (E03, E04, E07), and the need to contextualize the uses made of artificial intelligence and all technologies under that umbrella (machine learning, deep learning, big data, etc.), with a view to improving the data collection processes that AI uses to train itself to avoid the replication of biases, among other issues. This issue was identified by all the interviewees, especially due to the fact that AI is "trained" with data that are biased (E03) or offer an inaccurate reflection of reality (E04, E06). The result of applying these poorly trained algorithms is the exacerbation of the inequality between men and women (E07).

E01: Nothing is a panacea. (...). At present I wouldn't say all of them, but there are big data programs with biases, and what they do is hire women less. (...) If you simply repeat the algorithm and you've been hiring more men all your life, then the algorithm will recommend hiring more men.

Furthermore, an issue pointed out by most interviewees was related to the initiatives that institutions and civic associations involved in data sharing and open data need to implement to increase gender visibility (E03, E04, E07). This issue involves two dimensions: a consideration of the organizations that are currently involved in collecting and sharing data (e.g., Open Street Maps, WikiData, Data Beers, Hackathon Lovers), and continuing to create events that interest and motivate users (E06), sharing the

need to develop skills to understand and handle data (E07); and the recognition of female role models in each of the professional fields in which the interviewees work, an aspect considered vital by the respondents (E02, E04, E05).

Among the activities of this kind to raise awareness about the importance of data in contemporary society, there was widespread agreement on the important role played by journalism as an intermediary between those who generate open data and the general public (E03). Before other activities related to data literacy that are implemented in formal educational contexts, E01 highlights the “essential role” of journalists to present interesting data visualizations that enable society to interpret the data and help identify what the source of the data is, where it comes from, etc. The following observations were made by the interviewee, who has direct experience in the publishing sector:

E05: The distortion and misuse of data is widespread, and our responsibility lies in turning them into means or ways of doing journalism that is reliable enough for people to know where to turn when they need to counter or contrast the arguments of politicians, for example, but also of any type of social actor.

As reflected in interviews, there is still work to be done to improve communication between data generators and society.

Discussion and conclusion

The responses of the women interviewed in this study confirm the continued existence of gaps, biases, and stereotypes in the open data sector. There are thus implicit references to phenomena such as the “glass ceiling,” the “sticky floor,” and the “imposter syndrome.” Moreover, awareness of significant deficiencies in the STEM sector (Petzel & Casad, 2022; Verdugo-Castro et al., 2022) and doubts about the real power of women in management positions when their appointment is in response to a quota system (Hamplová et al., 2022) are issues that have marked their academic and professional development and their experience working with data.

These inequalities are corroborated in the existing literature on the gender gap in the digital sector (Eurostat, 2019), and the interviews conducted for this research confirm that this gap also exists in the open data sector. There are very few publications that address this phenomenon from the perspective of the social value of open data and how they can contribute to the design, approval, and implementation of effective policies for achieving real equality between women and men.

The fact that this is a new field closely associated with development and innovation has not prevented it from inheriting the same obstacles that hinder gender equality in other contexts. As the interviewees observed, technology that feeds on data is essentially designed from a male perspective, as it is a niche market occupied by more men than women, which has also been the case historically in other STEM fields. The respondents’ observations are corroborated by the fact that a relatively low percentage of women are working in jobs associated with technology in the European Union (less than 20% of ICT professionals and of ICT technicians), and that these professions have particularly low levels of unemployment (less than 5%), with both types of ICT employment (professionals and technicians) exceeding 100% of relative income compared to similar occupations (Cedefop, 2023).

In this regard, Leavy (2018) claims that to avoid gender-biased algorithms influencing decisions in our society, diversity in the area of machine learning is important. She argues that this diversity in the composition of groups of AI developers could be a way of addressing the issue of gender bias by better evaluating training data, incorporating equity concepts into algorithms, and assessing the potential impact of gender bias in the context of technological development. Specifically, she highlights the benefits of incorporating gender theory into machine learning, pointing to the example of textual data where machine learning algorithms that are trained on older corpora may reflect outdated ways of referring to men and women. Similarly, Manasi et al. (2022) stress the importance of including women in the technology sector and the harmful consequences of developing AI tools without a gender perspective.

Furthermore, in relation to the integration of a gender perspective into the open data sector, it has been found that the disaggregation of data continues to be a key challenge in all the sectors analyzed in this study (Grantham, 2020; Open Data Watch, 2022). In particular, the absence of disaggregated data has been identified in areas as diverse as journalism, data science, feminist activism, and public administration. This absence has influenced not only the existing imbalances in terms of availability between data considered hard (associated with “hard skills”) and soft (areas traditionally associated with women), but also the inadequate use made of certain data catalogs that are in fact disaggregated.

In relation to this imbalance, an interesting study by Gorska and Jemielniak (2023) examines the influence of gender bias in the AI-generated images of people in socially recognized professions using four different AI-generators. The study found a significant gender bias, with men represented in 76% of the images and women in only 8%.

The interviews also reveal that although data do not have a gender perspective *per se*, the lack of a gender perspective in government agencies, scientific research, and industry, especially in companies in the technology sector, will result in the collection of data that fails to account for the range of variables related to sex and gender, and will subsequently make it difficult to process the data in a way that considers the particularities associated with these variables. It is important to remember that open data and the developments associated with them, such as AI systems, are a product of human creation and they therefore inherit their developers’ way of seeing the world; thus, the popular belief that data are intrinsically objective is fundamentally flawed (Gorska & Jemielniak, 2023).

The solution, highlighted by all the interviewees without exception, involves the development of interoperable and consistent data collection and analysis models that integrate a gender perspective right from the outset, as well as the training of professionals in fields associated with the use of open data.

Central to this process is the role of government agencies, which should take decisions that encourage the integration of women into educational contexts associated with digital and automated environments and promote efficient public policies that give greater visibility to the role of women in areas of knowledge associated with open data.

In the current context, the importance of AI and its associated technologies must inevitably be acknowledged. Based on the interviewees’ responses, an opportunity exists that has yet to be fully seized upon. To do so, the data used to train algorithms should be collected with a gender perspective, as only in this way can the replication of biases

that harm a significant proportion of the population be avoided (Bolukbasi et al., 2016; Castillo, 2023; Wang et al., 2019). To meet this challenge, all actors need to be involved, including government agencies, academic institutions, and the private sector, in a task that largely involves promoting diversity on work teams, and contributing different perspectives that include different, traditionally excluded groups right from the outset.

Recovering concrete examples reported by the interviewees, the initiatives that institutions and civic associations involved in data sharing and open data such as Open Street Maps, WikiData, Data Beers, Hackathon Lovers need to increase gender visibility and the recognition of female role models. This is even more relevant if the final product is related to issues that specifically affect women's lives, for instance (in the words of the interviewees) if we want to make an app about menstruation, we should include women in the development team; or if we want to design an app about dating, the specific characteristics of female users should be taken into account. In this sense, it is worth remembering that there are already women's groups in the field of AI such as R Ladies, PyLadies, #Wethewomenjournalistsstop or the Chicas Poderosas network in Latin America, which are working very intensively on these issues. Also, other projects such as *Data2x*, *Digital fems* or *GenderDataLab* have been created to raise awareness and visibility to open data and gender perspective, whereas others such as *APP anticipate*, *Gender Data Kit* or *Medicamentalia Anticonceptivos* are focused on how open data can help prevent gender-based violence (Sixto-Costoya & Ferrer-Sapena, 2023).

Considering both the biases and shortcomings as well as the initiatives carried out to incorporate a gender perspective in open data, it can be said that we are moving at different speeds (Sixto-Costoya & Ferrer-Sapena, 2023). For this reason, it is important to highlight the need to work on making the public aware of the importance of data, encouraging them to take part in the valuable role that data play, not merely as passive data generators, but as users and developers, empowering them to take possession of this new (or now not so new) vital resource that has become a part of our lives.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Lorena R. Romero-Domínguez, BA and PhD in Journalism. Member of the scientific researcher group "Media, media policy and democracy". Currently she is Associate Professor in the Department of Journalism at the University of Sevilla, Spain. She was visiting professor in the following Universities: Johannes-Gutenberg Mainz Universität (Germany), Cardiff University (Wales), Konstanz Universität (Germany), Humboldt Universität zu Berlin (Germany). She is Co-editor of the academic collection "Communication Sciences" (Editorial Universidad de Sevilla), member of the Observatory of Transparency and Open Data (Polytechnic University of Valencia) and evaluator of the Quality Assurance Agency for the University System in Castilla y Leon and for the Canarian University System. Her main research areas include journalism; Cross-border Investigative Journalism; Media and Communication History; Media Economics, Policy and Regulation; and Open Data and Journalism.

Andrea Sixto-Costoya, BA in Social Work; Master in Research Treatment and Associated Pathologies in Drug Addictions; and PhD from the University of Valencia. She is currently

Assistant Professor at the Department of Social Work at the University of Valencia. Throughout her career she has obtained prestigious competitive predoctoral and postdoctoral fellowships, and has carried out different research stays both nationally and internationally. She is part of the Research and Social and Health Information Unit (UISYS group) of the University of Valencia and is a member of the Observatory of Transparency and Open Data (Polytechnic University of Valencia). She has published several articles in journals of international prestige, as well as book chapters. She has participated in competitive research projects related to the areas of addictions, gender perspective in health sciences and open science and open data. At the transfer level, she has been part of agreements with the City Council of Valencia, the Department of Health of the Generalitat Valenciana; and the Department of Participation, Transparency, Cooperation and Democratic Quality of the Generalitat Valenciana.

Antonia Ferrer-Sapena, BA in Geography and History, and PhD in Information Sciences. She is currently Professor at the Universitat Politècnica de València. She is chair of the Madphy research group at the Polytechnic University of Valencia and director of the Chair and Observatory of Open Data and Transparency. She is an advisory member of La Ciba (Women's resource space, innovation and feminist economics), member of the participation group of the Valencian Anti-Fraud Agency (AVA), and vice-president of the Tactical Whistleblowers Association. She has directed more than 60 Final Degree Projects and Master's Theses. Her main lines of research currently focus on open access to scientific knowledge and open research data. She has participated in 30 R&D projects of which she has been Principal Investigator in 7 of them. She has more than 20 participations in contracts, agreements or non-competitive R+D+i projects with public or private administrations or entities. In terms of publications, she has published more than 60 publications in journals and book chapters.

Funding

This publication has been produced with the financial support of the Conselleria de Participación, Transparencia, Cooperación y Calidad Democrática. The contents of this publication are the sole responsibility of the Universitat Politècnica de València and do not necessarily represent the Generalitat Valenciana. The authors received the ASC postdoctoral grant from the Spanish Ministry of Education and Professional Training (MS21-020).

ORCID

Lorena R. Romero-Domínguez  <http://orcid.org/0000-0001-6942-0183>

Andrea Sixto-Costoya  <http://orcid.org/0000-0001-9162-8992>

Antonia Ferrer-Sapena  <http://orcid.org/0000-0001-6432-917X>

Data availability statement

The raw data generated for this study were deposited in the Zenodo repository and can be consulted through the following link: <https://doi.org/10.5281/zenodo.10698522>.

References

- Álvarez-Díaz, J. A. (2020). The necessary gender perspective for the analysis of health problems. *Cirugía y Cirujanos*, 88(3), 383–388. <https://doi.org/10.24875/CIRU.19000865>
- Barr, E., & Temkin, S. (2022). *Gender as a social and cultural variable and health*. National Institutes of Health. <https://orwh.od.nih.gov/about/director/messages/gender-as-social-and-cultural-variable-and-health>
- Black, C. F. (2016). Global gender gap report. In *Encyclopedia of Family Studies*. <https://doi.org/10.1002/9781119085621.wbefs350>

- Bolukbasi, T., Chang, K. W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. arXiv. <https://doi.org/10.48550/arXiv.1607.06520>
- Caetano, A., & Moran, L. (2021). Biographical research through the looking glass of pandemic times: Challenges and possibilities of online interviewing. *Irish Journal of Sociology*, 30(2), 209–213. <https://doi.org/10.1177/07916035211022182>
- Castillo, D. (2023). *The gender data gap in AI: Confronting bias in machine learning*. <https://www.seldon.io/the-gender-data-gap-in-ai>
- Cedefop. (2023). *Browse by occupation*. <https://www.cedefop.europa.eu/en/tools/skills-intelligence/occupations?occupation=3.35#2>
- Cho, A. H., Johnson, S. A., Schuman, C. E., Adler, J. M., Gonzalez, O., Graves, S. J., Huebner, J. R., Blaine Marchant, D., Rifai, S. W., Skinner, I., & Bruna, E. M. (2014). Women are underrepresented on the editorial boards of journals in environmental biology and natural resource management. *PeerJ*, 2(1), e542. <https://doi.org/10.7717/peerj.542>
- Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., Gigante, A., Valencia, A., Rementeria, M. J., Chadha, A. S., & Mavridis, N. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digital Medicine*, 3(1), 81. <https://doi.org/10.1038/s41746-020-0288-5>
- Comisión Europea. (2020). *Comunicación de la comisión al parlamento europeo, al consejo, al comité económico y social europeo y al comité de las regiones*. Una Unión de la igualdad: Estrategia para la Igualdad de Género 2020-2025. <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0152>
- Corbetta, P. (2007). Metodología y técnicas de investigación social [Methodology and techniques of social research]. McGraw-Hill Interamericana de España.
- Creswell, J. (2012). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Sage.
- Curasi, C. F. (2001). A critical exploration of face-to face interviewing vs. computer mediated interviewing. *International Journal of Market Research*, 43(4), 1–13. <https://doi.org/10.1177/147078530104300402>
- Díaz-Bravo, L., Torruco-García, U., Martínez-Hernández, M., & Varela-Ruiz, M. (2013). La entrevista, recurso flexible y dinámico [The interview, a flexible and dynamic resource]. *Investigación En Educación Médica*, 2(7), 162–167. [https://doi.org/10.1016/S2007-5057\(13\)72706-6](https://doi.org/10.1016/S2007-5057(13)72706-6)
- Dimond, J., Fiesler, C., DiSalvo, B., Pelc, J., & Bruckman, A. (2012, October 27–31). Qualitative data collection technologies: A comparison of instant messaging, email, and phone. GROUP'12. <http://jilldimond.com/wp-content/uploads/2010/10/group115-dimond.pdf>
- Domínguez Prats, P., & Cuesta, J. (2019). Women in academia from the 19th to the 21st century. *Culture & History Digital Journal*, 8(1), e001–e001.
- Edwards, R., & Holland, J. (2020). Reviewing challenges and the future for qualitative interviewing. *International Journal of Social Research Methodology*, 23(5), 581–592. <https://doi.org/10.1080/13645579.2020.1766767>
- Erren, T. C., Groß, J. V., Shaw, D. M., & Selle, B. (2014). Representation of women as authors, reviewers, editors in chief, and editorial board members at 6 general medical journals in 2010 and 2011. *JAMA Internal Medicine*, 174(4), 633–635. <https://doi.org/10.1001/jamainternmed.2013.14760>
- Escobar-Pérez, J., & Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización [Content validity and expert judgment: An approach to their use]. *Avances En Medición*, 6, 27–36. http://www.humanas.unal.edu.co/psicometria/files/7113/8574/5708/Articulo3_Juicio_d_e_expertos_27-36.pdf
- European Commission. (2021). *She figures 2021 gender in research and innovation: Statistics and indicators*. European Commission. <https://doi.org/10.2777/06090>
- European Commission. (2022). *Gender equality in the EU justice and consumers*. European Commission. <https://doi.org/10.2838/94579>
- Eurostat. (2019). *Women in science and technology*. <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20190211-1>
- Ferrer-Sapena, A., Peset, F., & Alexandre-Benavent, R. (2011). Acceso a los datos públicos y su reutilización: Open data y open government [Access to and reuse of public data: Open data

- and open government]. *Profesional de La Informacion*, 20(3), 260–269. <https://doi.org/10.3145/epi.2011.may.03>
- Gainza Veloso, A., & Canales Cerón, E. M. (2006). La entrevista en profundidad individual [The individual in-depth interview]. In *Metodologías de la investigación social* (pp. 219–263). LOM Ediciones.
- Gogos, A., Langmead, C., Sullivan, J. C., & Lawrence, A. J. (2019). The importance of sex differences in pharmacology research. *British Journal of Pharmacology*, 176(21), 4087–4089. <https://doi.org/10.1111/bph.14819>
- Gorska, A. M., & Jemielniak, D. (2023). The invisible women: uncovering gender bias in AI-generated images of professionals. *Feminist Media Studies*, 23(8), 4370–4375. <https://doi.org/10.1080/14680777.2023.2263659>
- Grantham, K. (2020). *Mapping gender data gaps: And SDG era update*. Data2x. <https://data2x.org/resource-center/mappinggenderdatagaps/>
- Gruber, M., Eberl, J. M., Lind, F., & Boomgaarden, H. G. (2020). Qualitative interviews with irregular migrants in times of COVID-19: Recourse to remote interview techniques as a possible methodological adjustment. *Forum Qualitative Sozialforschung Forum: Qualitative Social Research*, 22(1), 1–15. <https://doi.org/10.17169/fqs-22.1.3563>
- Hamplová, E., Janeček, V., & Lefley, F. (2022). Board gender diversity and women in leadership positions – Are quotas the solution? *Corporate Communications: An International Journal*, 27(4), 742–759. <https://doi.org/10.1108/CCIJ-02-2022-0022>
- Helbig, R., Von Höveling, S., Solsbach, A., & Gómez, J. M. (2021). Strategic analysis of providing corporate sustainability open data. *Intelligent Systems in Accounting, Finance and Management*, 28(3), 195–214. <https://doi.org/10.1002/isaf.1501>
- Herala, A., Kokkola, J., Kasurinen, J., & Vanhala, E. (2019). Strategy for data: Open it or hack it? *Journal of Theoretical and Applied Electronic Commerce Research*, 14(2), 33–46. <https://doi.org/10.4067/S0718-18762019000200104>
- Irani, E. (2019). The use of videoconferencing for qualitative interviewing: Opportunities, challenges, and considerations. *Clinical Nursing Research*, 28(1), 3–8. <https://doi.org/10.1177/1054773818803170>
- Janghorban, R., Latifnejad Roudsari, R., & Taghipour, A. (2014). Skype interviewing: The new generation of online synchronous interview in qualitative research. *International Journal of Qualitative Studies on Health and Well-Being*, 9(1), 24152. <https://doi.org/10.3402/qhw.v9.24152>
- Kelly, S. E. (2010). Qualitative interviewing techniques and styles. In I. Bourgeault, R. Dingwall, & R. De Vries (Eds.), *The SAGE handbook of qualitative methods in health research* (pp. 307–326). SAGE Publications Ltd. <https://doi.org/10.4135/9781446268247>
- Lancharro Taveró, I., Gil García, E., Macías Seda, J., Romero Serrano, R., Calvo Cabrera, I. M., & Arroyo Rodríguez, A. (2018). The gender perspective in the opinions and discourse of women about caregiving. *Revista Da Escola de Enfermagem Da U S P*, 52, e03370. <https://doi.org/10.1590/S1980-220X2017009403370>
- Leavy, S. (2018). Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning. In *Proceedings International Conference on Software Engineering* (pp. 14–16). <https://doi.org/10.1145/3195570.3195580>
- Lee, M. S., Guo, L. N., & Nambudiri, V. E. (2022). Towards gender equity in artificial intelligence and machine learning applications in dermatology. *Journal of the American Medical Informatics Association: JAMIA*, 29(2), 400–403. <https://doi.org/10.1093/jamia/ocab113>
- Lee, S. K. (2018). Sex as an important biological variable in biomedical research. *BMB Reports*, 51(4), 167–173. <https://doi.org/10.5483/BMBRep.2018.51.4.034>
- Lopezosa, C., Codina, L., & Freixa, P. (2022). *ATLAS.ti para entrevistas semiestructuradas: guía de uso para un análisis cualitativo eficaz*[ATLAS.ti for semi-structured interviews: A user's guide for effective qualitative analysis]. Serie Editorial DigiDoc.
- Manasi, A., Panchanadeswaran, S., Sours, E., & Lee, S. J. (2022). Mirroring the bias: Gender and artificial intelligence. *Gender, Technology and Development*, 26(3), 295–305. <https://doi.org/10.1080/09718524.2022.2128254>
- Meho, L. I. (2006). E-mail interviewing in qualitative research: A methodological discussion. *Journal of the American Society for Information Science and Technology*, 57(10), 1284–1295. <https://doi.org/10.1002/asi.20416>

- Nathan, S., Newman, C., & Lancaster, K. (2019). Qualitative interviewing. In P. Liamputtong (Ed.), *Handbook of research methods in health social sciences* (pp. 391–410). Springer. https://doi.org/10.1007/978-981-10-5251-4_77
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Sage.
- O'Connor, H., & Madge, C. (2001). Cyber-mothers: Online synchronous interviewing using conferencing software. *Sociological Research Online*, 5(4), 102–117. <https://doi.org/10.5153/sro.543>
- Olson, K. (2011). *Essentials of qualitative interviewing*. Routledge.
- OpenCorporate. (2023). *Our purpose*. <https://opencorporates.com/info/about/>
- Open Data Watch. (2021). Open data inventory. Data2x. <https://opendatawatch.com/publications/odin-open-data-inventory-2020-21-executive-summary/>
- Open Data Watch. (2022). Transforming the data landscape: Solutions to close gender data gaps. Data2x. <https://data2x.org/resource-center/transforming-the-data-landscape-solutions-to-close-gender-data-gaps/>
- Otterbacher, J., Bates, J., & Cloug, P. (2017). Competent men and warm women: Gender stereotypes and backlash in image search results. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 6620–6631). Association for Computing Machinery. <https://doi.org/10.1145/3025453.3025727>
- Petzel, Z. W., & Casad, B. J. (2022). Take a chance on STEM: Risk-taking buffers negative effects of stereotype threat. *The Journal of Experimental Education*, 90(3), 656–672. <https://doi.org/10.1080/00220973.2020.1848766>
- Prages. (2009). *Guidelines for gender equality programmes in science*. European Commission.
- Romero Domínguez, L. R., Sixto-Costoya, A., & Ferrer-Sapena, A. (2024). *Interview transcripts for the study entitled "Exploring stereotypes, bias, and expectations of women in the open data context" (Versión 1) [Data set]*. Zenodo. <https://doi.org/10.5281/zenodo.10698522>
- Ruggieri, R., Pecoraro, F., & Luzi, D. (2021). An intersectional approach to analyse gender productivity and open access: A bibliometric analysis of the Italian National Research Council. *Scientometrics*, 126(2), 1647–1673. <https://doi.org/10.1007/s11192-020-03802-0>
- Sixto-Costoya, A., Alonso-Arroyo, A., Castelló-Cogollo, L., Aleixandre-Benavent, R., & Valderrama-Zurián, J. C. (2022). Gender presence on the editorial boards of journals in the Women's Studies subject category. *Women's Studies International Forum*, 93(June). <https://doi.org/10.1016/j.wsif.2022.102617>
- Sixto-Costoya, A., & Ferrer-Sapena, A. (2023). Datos abiertos y perspectiva de género, ¿por dónde vamos? *Anuario ThinkEPI*, 17, 1–8. <https://doi.org/10.3145/thinkepi.2023.e17a33>
- Sociedad Max Planck. (2003). La Declaración de Berlín sobre acceso abierto. *GeoTropico*, 1(2), 152–154.
- Spector, N. D., & Overholser, B. (2019). Examining gender disparity in medicine and setting a course forward. *JAMA Network Open*, 2(6), e196484. <https://doi.org/10.1001/jamanetworkopen.2019.6484>
- Tandon, A., & Pareek, V. (2019). Gender data gap. In A. Tandon, T. Mewa, & S. Chattapadhyay (Eds.), *Handbook of gender and open data*. <https://cis.pubpub.org/pub/gender-data-gap/release/6>
- The World Bank. (2023). *Gender data portal*. <https://genderdata.worldbank.org/>
- Tomassini, C. (2021). Brechas de género en la ciencia: revisión sistemática de las principales explicaciones y agenda de investigación [Gender gaps in science: Systematic review of the main explanations and research agenda]. *Education in the Knowledge Society (EKS)*, 22. <https://doi.org/10.14201/eks.25437>
- Tonon, G. (2012). Reflexiones latinoamericanas sobre investigación cualitativa [Latin American reflections on qualitative research]. *Revista Latinoamericana de Ciencias Sociales, Niñez Y Juventud*, 8(1). <https://revistaumanizales.cinde.org.co/rlnsj/index.php/Revista-Latinoamericana/article/view/606>
- UNESCO. (2014). *Plan de Acción de la UNESCO para la prioridad Igualdad de Género (2014-2021)*. https://unesdoc.unesco.org/ark:/48223/pf0000227222_spa?posInSet=6&queryId=be8e8136-a047-40cd-9402-b072729bb296
- United Nations. (2016). *Integrating a gender perspective into statistics*. <https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Handbooks/gender/Integrating-a-Gender-Perspective-into-Statistics-E.pdf>

- United Nations. (2022a). *Full and equal access and participation for women and girls in science*. <https://www.un.org/en/observances/women-and-girls-in-science-day>
- United Nations. (2022b). *The sustainable development goals report 2019*. United Nations publication issued by the Department of Economic and Social Affairs. <https://unstats.un.org/sdgs/report/2022/The-Sustainable-Development-Goals-Report-2022.pdf>
- Verdugo-Castro, S., García-Holgado, A., & Sánchez-Gómez, M. C. (2022). The gender gap in higher STEM studies: A systematic literature review. *Heliyon*, 8(8), e10300. <https://doi.org/10.1016/j.heliyon.2022.e10300>
- Wang, T., Zhao, J., Yatskar, M., Chang, K. W., & Ordonez, V. (2019). Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations. arXiv. <https://doi.org/10.48550/arXiv.1811.08489>
- World Health Organization. (2018). *Gender and Health*. <https://www.who.int/es/news-room/fact-sheets/detail/gender>