



HAL
open science

Intersectional Study of the Gender Gap in STEM through the Identification of Missing Datasets about Women: A Multisided Problem

Genoveva Vargas-Solar

► **To cite this version:**

Genoveva Vargas-Solar. Intersectional Study of the Gender Gap in STEM through the Identification of Missing Datasets about Women: A Multisided Problem. Applied Sciences, 2022, 12 (12), 10.3390/app12125813 . hal-03691088

HAL Id: hal-03691088

<https://hal.science/hal-03691088v1>

Submitted on 8 Jun 2022

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

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

Article

Intersectional Study of the Gender Gap in STEM through the Identification of Missing Datasets about Women: A Multisided Problem

Genoveva Vargas-Solar 

CNRS, University Lyon, INSA Lyon, UCBL, Ecole Centrale Lyon, University Lyon 2, LIRIS, UMR5205, 69622 Villeurbanne, France; genoveva.vargas-solar@cnrs.fr

Abstract: This paper discusses the problem of missing datasets for analysing and exhibiting the role of women in STEM with a particular focus on computer science (CS), artificial intelligence (AI) and data science (DS). It discusses the problem in a concrete case of a global south country (i.e., Mexico). Our study aims to point out missing datasets to identify invisible information regarding women and the implications when studying the gender gap in different STEM disciplines. Missing datasets about women in STEM show that the first step to understanding gender imbalance in STEM is building women's history by "completing" existing datasets.

Keywords: data; missing datasets; gender gap; data-driven studies; women in artificial intelligence; women in data science; women in STEM



Citation: Vargas-Solar, G.

Intersectional Study of the Gender Gap in STEM through the Identification of Missing Datasets about Women: A Multisided Problem. *Appl. Sci.* **2022**, *12*, 5813. <https://doi.org/10.3390/app12125813>

Academic Editors: Aida Valls and Karina Gibert

Received: 1 April 2022

Accepted: 2 June 2022

Published: 8 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. 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/>).

1. Introduction

Many science and engineering occupations are predicted to grow faster than the average rate for all professions. Workforce projections in 2018 by the U.S. Department of Labour showed that nine of the ten fastest-growing occupations that require at least a bachelor's degree would need significant scientific or mathematical training [1]. This prediction has been confirmed in recent years, and computer science fields with great potential like artificial intelligence (AI) and data science (DS) can be certainly included. More recent sources anticipate that 75% of future jobs will be related to these fields (<https://lac.unwomen.org/en/digiteca/publicaciones/2020/09/mujeres-en-ciencia-tecnologia-ingenieria-y-matematicas-en-america-latina-y-el-caribe>, accessed on 1 June 2022), given that 7.1 million jobs were expected to be displaced by 2020, and half of existing jobs will disappear by 2050. Some of the most significant increases will be in engineering (and computer-related) fields in which women currently hold one-quarter or fewer positions [2,3]. Indeed, only 22% of all professionals working in AI around the world are women (World Economic Forum. Global Gender Gap Report 2018, <https://www.weforum.org/reports/the-global-gender-gap-report-2018/>, accessed on 1 June 2022). The ubiquitous male dominance in countries in different regions results in a feedback loop shaping gender bias in AI and machine learning systems used in DS experiments [4–6]. These fields are particularly fast-moving [7] both in industry and academia, so it is essential to map how gender gaps (Davenport, T.H. and Patil, D.J. (2012) Data Scientist: The Sexiest Job of the 21st Century. Harvard Business Review. Retrieved from: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>, accessed on 1 June 2022) [8] are manifest in data-driven solutions comprehensively. If not addressed soon, the gender gap in STEM will widen during the Fourth Industrial Revolution. Science and engineering address key challenges of our time:

- Finding cures for diseases like cancer and malaria, tackling global warming, providing people with clean drinking water, developing renewable energy sources, and understanding the origins of the universe;

- Engineers design many of the things we use daily, including buildings, bridges, computers, cars, wheelchairs, and X-ray machines.

The presence of women in teams seem to increase collective intelligence, and if those women have decision making positions, success probabilities increase in startups. When women are not involved in designing products and addressing social and political problems, then needs and desires unique to women may be overlooked. For example, a predominantly male group of engineers tailored the first generation of automotive airbags to adult male bodies, resulting in avoidable deaths for women and children [9]. With a more diverse workforce, scientific and technological products, services, and solutions are likely to be better designed and more likely to represent all users.

The consolidation of AI and DS (World Economic Forum. Global Gender Gap Report 2018. Retrieved from: <https://www.weforum.org/reports/the-global-gender-gap-report-2018>, accessed on 1 June 2022, World Economic Forum. Global Gender Gap Report 2020. Retrieved from: <https://www.weforum.org/reports/gender-gap-2020-report-100-years-pay-equality>, accessed on 1 June 2022, World Economic Forum. The Future of Gender Parity: A Labour Market Shift. Retrieved from: https://reports.weforum.org/global-gender-gap-report-2020/the-future-of-gender-parity/?doing_wp_cron=1612906910.1824119091033935546875, accessed on 1 June 2022, as backbone tools to promote data-driven solutions opens the risk of overlooking women in the loop. Women are underrepresented in observation and data collection tasks. Their workforce in AI and DS is small, and they are underrepresented in algorithm design and testing teams. Thus, their characteristics are not represented because of how data are engineered. Their needs are not visible for designing units, even if more than the 50% of the planet's population are women (The National Centre for Women & Information Technology (NCWIT <https://ncwit.org>, accessed on 1 June 2022.) argues about the importance and interest for organizations to search for gender and race equality. Besides the economic welfare, teams respecting gender equality seem to be more creative and effective in performing tasks and sharing knowledge).

When science is combined with economy, politics and social impact, it is possible to classify disciplines and reason about the capital they produce and who produces that capital. To activate the economy, every asset that is potentially productive must participate in the active STEM market. Women are a massive part of the assets that remain excluded from the equation, weakening the economy. Countries with dynamic economies like Germany lack human force in these disciplines and import brains to fulfil the requirements of their economy. According to the economist François Lenglet in his TV show "La guerre des âges", two strategies are possible for these economies: either hiring an immigrant brain force or integrating women into the production market. Surprisingly, Germany is developing a very aggressive strategy for attracting immigrant experts (!) (Is the old German saying that associates women to the three K's (i.e., Kirche, Küche, Kinder), still valid? The German gender issues are out of the scope of this paper. Thus, we leave this topic open and we invite the interested reader willing to seek for more information to visit [10,11]). Why are women excluded from the equation? Are they recognised as a potential labour force, or are they hidden Cinderellas invisible to statistics? Are gender bias and discrimination playing a role in women exclusion in STEM?

We believe that it is critical to study the phenomenon of data and women's absence in STEM from many different perspectives. Since women come in many different sizes, colours, and formats, the study must consider several communities from the global north and south instead of general studies.

This paper studies missing data under a qualitative approach [12] that combines grounded theory and use case strategies for answering the following research questions:

- RQ1 Do missing data concerning women's STEM research participation prevent a comprehensive view of their contribution across history?
- RQ2 Is it possible to design viable data science experiments that can estimate the female workforce in STEM and answer the question, "where are women in STEM"?

RQ3 Do missing intersectional perspectives prevent performing representative quantitative analysis about women's activity and contribution in STEM?

Our work adopts a qualitative methodology that does not rely upon mathematical, data mining or artificial intelligence models. Still, this "alternative" analytics choice can be seen as a preparation protocol for data science experiments that use AI models because of the critical role of data in ensuring "representative" results. These models are adapted in cases with datasets that are representative observation samples of a phenomenon, even if raw. Yet, the hypothesis that drives this study is that missing datasets giving insight into female contribution are missing. We question the possibility of producing statistical and induced or deduced knowledge from incomplete observations.

1.1. Context and Methodology

For performing a study on missing data and female absence as two perspectives of a mirror, it is essential to consider the conditions in which data were absent before and after the pandemic years. In the global north, there are programs like UN Women (<https://lac.unwomen.org/en>, accessed on 1 June 2022), She Figures (<https://ec.europa.eu/assets/rtd/shefigures2021/index.html>, accessed on 1 June 2022), and Davos reports (https://www.eversheds-sutherland.com/global/en/what/articles/index.page?ArticleID=en/Data-Protection/World_Economic_Forum_Report_2022, accessed on 1 June 2022) that have collected data about women in the economy, particularly in STEM. They perform serious data collection, and they often rely on governmental data. Other works like [13] perform analytics processes to measure the gender gap index in schools (before university and possible activities that can generate scientific contribution) by combining qualitative and quantitative methods. However, data are not intersectional (Intersectionality is an analytical framework for understanding how aspects of a person's social and political identities converge to create different modes of discrimination and privilege, <https://en.wikipedia.org/wiki/Intersectionality>, accessed on 1 June 2022); companies and institutions do not share them, so they are not included in studies. In consequence, despite these efforts, data about women are still missing. Thus, it is capital to understand the forces that play in the persistence of this discriminatory situation and the distribution of opportunities [14]. We need datasets as complete as possible, and as fine-grained and representative as possible, to perform clear data-driven studies about gender balance. Yet, these studies' fuel (i.e., datasets) is partial, incomplete or absent. It is not a matter of the analytics processes (data preparation, cleaning, engineering) nor the analytics models applied to extract knowledge. The point is that without intersectional data, which is often missing in available datasets, the gender issue cannot be modelled and understood in its complexity. These studies can lead to conclusions but do not wholly understand the gender gap problems. Yet, to efficiently address gender gap problems, we believe that we need to identify which datasets are missing. Our work focuses on tracking data that provide evidence about women's participation in STEM, particularly in CS, AI and DS.

Therefore, our work focusses on missing datasets instead of analysing the gender gap with available ones. These studies have been already done at the national level in different countries (e.g., in the UK [15], or specific datasets [16]) by international bodies like UN Women, the European Commission with the "She Figures" report, and by editorial bodies. These studies remain partial because they concentrate on quantitative approaches with datasets of one type (i.e., publications, student numbers); they lack intersectional approaches for collecting data [17,18].

Our first strategy is to identify women's influential role and contribution to the global north (The concept of Global North and Global South is used to describe a grouping of countries along socio-economic and political characteristics. The Global South identifies the regions of Latin America, Asia, Africa, and Oceania. It denotes regions outside Europe and North America, mostly (though not all) low-income and often politically or culturally marginalised countries on one side of the so-called divide, the other side being the countries of the Global North (often equated with developed countries). The term does not inherently

refer to a geographical south) in CS represented in the annals of relevant awards in the field. We have performed an exploratory study of the sources where data lies that is explicit or hidden about women in STEM.

1.2. Contributions

As a result of the grounded theory and use case based approach:

1. We enumerate and characterise datasets regarding women in DS and IA, which are timidly emerging;
2. We report on a data analytics use case in Mexico with information about the National System of Researchers, the Mexican Academy of Sciences, the Mexican National Award of Sciences and elements of the history of Mexican women in STEM.

The choice of Mexico as a case study is because our vision is to perform regional studies about women, mainly located in the global south.

Accordingly, the remainder of the paper is organised as follows. Section 2 introduces existing projects and initiatives devoted to building the history of women in CS with a particular interest in works addressing their role in AI and DS. It gives a timeline of women that have contributed to pin milestones in the history of CS. Section 3 describes a use case about Mexico highlighting possible sources where data refer to female scientific and technological contributions. Section 4 concludes the paper and discusses open issues and future work.

2. Making Women's Contribution in Computer Science Visible

Studying and understanding the gender gap in STEM is a complex problem because each discipline and scientific community has different characteristics and histories. To analyse missing datasets about the gender gap in STEM and identify the type of datasets missing, we decided to focus on CS and two related sub-disciplines, AI and DS. However, at some point, we refer to other disciplines, which shows to which extent data are missing that we need to look at the datasets that we have at hand and derive guesses about related fields.

Figure 1 shows the workflow of our qualitative approach and how it is related to the steps of a data science workflow. Our approach opens the data collection box and exhibits three main phases, namely, (1) selecting and choosing data providers and (2) building datasets both systematically interacting with a (3) qualitative assessment phase. In our approach, the first phase applies the grounded theory methodology to identify providers and the type and characteristics of provided data. It drives conclusions about them, launching the qualitative assessment phase. Once data providers have been selected, datasets are built by applying fusion, integration and correlation to the data stemming from different providers. The objective of this phase is to “complete” data representative for answering a research question that drives a data science experiment. Recall that data science uses quantitative methodology implemented by applying data mining, machine learning and other artificial intelligence techniques.

Qualitative phases of data science and data-driven studies are mainly done manually or with an active presence of a human in the loop. This characteristic explains why this study does not report the application of (semi)-automatic techniques for performing analytics. Still, we argue that qualitative methodologies preliminary to the automatic phases of a data science workflow are critical because they show the interaction between DM, ML and AI and human tasks.

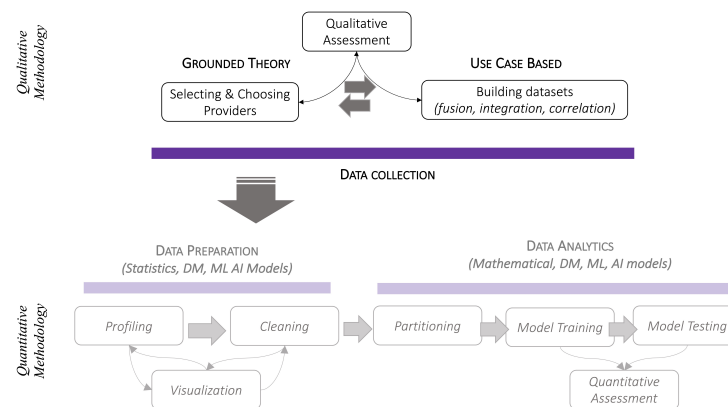


Figure 1. Qualitative methodology for identifying missing datasets.

We instantiated this approach in our study as follows. For selecting and choosing data providers (phase 1), we adopted the hypothesis that female contribution (from the global north and south) in STEM should be visible through the data representing their work and contribution. Data should refer to actions intended to create women collectivities in the area, the female workforce in promising topics like AI and DS, and prestigious awards. Research questions (RQ1, RQ2, RQ3) guide our grounded theory strategy, and it leads to conclusions about the qualitative characteristics of available datasets presented in the following sections. Similarly, the building datasets phase is driven by a use case strategy. Therefore, we report on observations and datasets available in Mexico that can compose a representative dataset about female contribution in STEM, particularly in computer science (see Section 3).

2.1. Women in STEM Studies and Inclusion Initiatives

The exclusion of the contribution of women scientists in STEM history is a concern in specific academic and industrial sectors around the world. Some studies suspect that the lack of critical mass of visible female scientists and professionals is one of the reasons for the increasing desertification of women in the field over the last three to four decades.

Different academic organisations in the global north like IEEE (women in engineering), ACM (women in ACM) in the US, the European Institute of Gender Equality in Europe, the CNRS through the “Mission pour la place des Femmes” and the “Comité parité-égalité”, the Institute of Gender in France, and major technology companies (e.g., Microsoft, Google, Facebook) have recognised the importance of understanding and organising actions that can promote a more gender-balanced, diverse and inclusive STEM (science, technology, engineering and mathematics) community. The opportunities and economic share, and return of investment in different disciplines in STEM are not homogeneous.

Some international actions in industry and academia have been organised to make women’s contributions to computing visible. One of the most established actions is the Anita Borg Institute, founded by computer science PhDs Anita Borg and Telle Whitney to recruit, retain and advance women in technology. Other international organisations are working to promote the visibility of women in computing:

- the Association for Computing Machinery (ACM) Committee on Women;
- the Association for Women in Computing;
- the Center for Women in Technology;
- Girl Develop It;
- Girls Who Code;
- In Latin America and the Caribbean, there is Meninas na Computação and Meninas Digitais of the Brazilian Computer Society; Mexicanas en Computació, of the Mexican Academy of Computing, the University of Chile organises the Latinity conference “Latinas in Computing”; the conference CLEI organises a special call for women in STEM in Latin America;

- Similar organisations in Africa, the Middle East, Asia-Pacific and Oceania have emerged.

This list is not representative of the diversity of groups and forums working for the inclusion and visibility of the contribution of women in computing. There are too many, and there is no integrated map enumerating them. This diversity demonstrates both the interest in the issue, the lack of coordinated actions, and the lack of scientific gender studies applied to women in computing and women in STEM considering fine-grained analysis, for example, by field or geographic region. Some governments and universities have created gender studies institutes with dedicated chairs for science, technology, engineering and mathematics. Yet, despite exceptions, the topic remains considered of second class and not an absolute priority.

Missing datasets 1: Who is studying and making women's work and contribution in STEM visible? In the datification era, the gender imbalance issue, women's contribution in raw datasets and representative samples for statistics are missing. For example, who is measuring the role of women in professional social networks, in DS forums like Kaggle, in technical discussions on Stack Overflow? The "absence" of (integrated) datasets and sources about women in STEM studies is a big issue. Without intersectional datasets, it is difficult to respond to the initial question: Where are women in STEM? What are the topics and problems they are addressing? How are they contributing to the economy of STEM? Who is collecting evidence and promoting studies to answer these questions?

2.2. Women in Artificial Intelligence and Data Science

The persistent absence of women employed in the AI and DS fields is troubling. According to the report of the World Economic Forum in 2018 (<https://www.weforum.org/reports/the-global-gender-gap-report-2018>, accessed on 1 June 2022), over three-quarters of professionals in these fields globally are male (78%); less than a quarter are women (22%). What about other underrepresented communities [19]? How are they represented in the DS and AI workforces, and which are the part of opportunities offered by these promising areas taken by these communities? Of course, to acquire a complete understanding of this phenomenon, it is necessary to treat the female community [1] and other underrepresented communities [19,20] as multifaceted and heterogeneous groups, with a plurality of experiences, and where gender intersects with multiple aspects of difference and disadvantage [21]. Discrimination at work must indeed be studied with an intersectional approach to acquire better and fine-grained understanding of the problem [22].

In the last decade, the role of data and scientific and technical skills used to exploit it have created promising career and economic spaces. AI and DS have emerged as promising areas for developing careers with critical financial benefit perspectives. Nevertheless, professional career perspectives in these disciplines are not equal depending on gender [23,24] and other criteria [25], including race/ethnicity, socio-economic level, the institution's reputation where people did their studies, etc. For example, in the United Kingdom, the House of Lords Select Committee on Artificial Intelligence in 2018 advocated increasing gender and ethnic diversity amongst AI developers. In France, several companies like Renault and Engie, through the "Laboratoire de l'Égalité", signed a call for widespread awareness of the discriminatory effects of AI and a commitment by its supporters to correct them. It is addressed to leaders in the public and private sectors, research and training organisations, companies that produce digital technology, companies that use digital technology and AI consultants. In 2020, the European Commission (European Commission (2019). 'Women in Digital Scoreboard'. Retrieved from: <https://digital-strategy.ec.europa.eu/en/library/women-digital-scoreboard-2020>, accessed on 1 June 2022. European Commission (2020a). Opinion on Artificial Intelligence—opportunities and challenges for gender equality. Advisory Committee on Equal Opportunities for Women and Men. (18 March). European Commission (2020b). Gendered Innovations 2: How Inclusive Analysis) noted that it is time to reflect on the interplay between AI and gender equality. French, European, and international organisations and agencies [8,26,27] perform studies for observing workforce shares in industry and some-

times in academia from a global perspective. Few fine-grained studies have studied underrepresented communities' workforce evolution and gaps from the gender perspective in disciplines such as AI and DS [6,28].

A thorough understanding of the way the workforce accesses the AI and DS opportunities in industry [29] and academia [30,31] is essential for building fair and inclusive societies [32]. This understanding can also be crucial for ensuring that countries obtain the full benefits of developing these areas to achieve better economic and social conditions and leading positions in the international arena through technology self-sufficiency.

Despite the economic and symbolic capital investment seeking a fair distribution of AI and DS opportunities for women and underrepresented communities, the crystal ceiling must still be broken [33]. Part of the explanation resides in data (!) [34]. Indeed, many studies agree to consider that quality, disaggregated, intersectional data are still missing. These data are essential to interrogate and tackle inequities in the AI and data science labour force [6]. As stated in the Alan Turing Institute study, "Where are women?" [6], the Royal Society has noted that a significant barrier to diversity is the lack of access to data on diversity statistics. The AI Roadmap recognises diversity and inclusion as a priority to make data-driven decisions to determine where to invest and ensure that underrepresented groups are given equal opportunity.

Missing datasets 2: Intersectional data about AI and DS female labour force. The existing evidence base about gender diversity in the AI and DS workforce is minimal. The available data is fragmented, incomplete and inadequate for investigating the career trajectories of women and men in the fields. Public data sets often rely upon data produced through proprietary analyses and methodologies. Governmental statistics lack detailed information about job titles and pay levels within ICT, computing, and technology. This partial vision of the workforce status is a significant barrier to examining the emerging hierarchy between data science, AI, and other subdomains. Furthermore, available data about the global AI & DS workforce is often aggregated and rarely broken down by age, race, geography, (dis)ability, sexual orientation, socioeconomic status, and gender. As stated by [6,35,36], "this is particularly concerning since it is those at the intersections of multiple marginalised groups who are at the greatest risk of being discriminated against at work and by resulting AI bias".

Integrating intersectional datasets allows understanding discrimination from different perspectives and mainly exhibits the various aspects that contribute to such discrimination. The absence of datasets about the female labour force in AI and DS is a form of discrimination. We believe it is necessary to promote actions applying mathematical and machine learning techniques to integrate complete and high quality, privacy-preserving data collections. Studies can use these data collections to drive conclusions about the gender gap in these disciplines.

2.3. *Awarded Women in Computer Science*

The history of computing seems to have an equivalent gender balance issue [9], as it has acknowledged with difficulty and marginally the contributions of women or at least their participation in the advances of this young science. Few documents outline the history of computer science, including women, and documents systematically include men: Alan Turing, Charles Babbage, Herman Hollerith, etc. The documents that list female computer scientists include Ada Lovelace, Hedy Lamarr, the ENIAC programmers (although their names and faces are unknown), Grace Hooper, Mary Allen Wikes, Lois Haibt or Radia Perlman. However, even members of the computing community are probably unaware of who these women scientists were and what their contributions were besides Ada Lovelace and Grace Hooper (the Appendix A lists a non-exhaustive but a more extended set of contributions authored by female computer scientists).

The history of women's contribution to computing is spread across blogs, websites and news articles. Some films acknowledge women's role as "calculators" [9,37–39], and female labour is mentioned as a curiosity in the history of science, insisting on "pencil-dragging"

tasks rather than on how they developed as programmers and became part of the core of digital computing advances [40]. The objective in referring to the term “calculator” is intended to show that contribution of women has been considered paper dragging even if they played a relevant role in the projects they participated in. Despite the importance of their contribution, they were not regarded as leaders of projects, and they remained invisible for a long time for history. The films have contributed to give visibility to these women. Still their stories and their contributions must be studied with methodological approaches and then included in books and in study programs, etc.

Histories of computing are not abundant, but they do exist; for example [41,42]. Reconstructing history is a complicated undertaking, to the extent that there are special series on the subject in well-known publishers such as Springer and IEEE. In Springer, the series is entitled “History of Computing” and in IEEE, it is entitled Annals of the History of Computing. These papers name the works of the illustrious scientists who have become the pillars on which computing is founded. Not surprisingly, the most prestigious prize in the field, the Turing Award, has only been awarded to three female scientists since it was first awarded in 1966. This event first occurred in 2006 (40 years later!). Can you name three or five names of people who have received the award? Are there any women’s names on your list? The names of the three female Turing Award laureates are:

- Frances E. Allen, pioneer in compiler optimisation, IBM Emeritus (2006);
- Barbara Liskov, programming languages, operating systems and innovations that have led to data abstraction, modularity, fault tolerance, persistence and distributed computing, Massachusetts Institute of Technology (MIT, 2008);
- Shafi Goldwasser, complexity theory, cryptography and number theory, MIT and Weizmann Institute of Science (2012).

The cases of other awards are similar. For example, the ACM SIGMOD Contributions Award in the area of databases, initiated in 1992, has recognised the work of Maria Zmankova (1992), Laura Haas (2000 with Michael Carey), Marianne Wenslett (2012) and Meral Özsoyoğlu (2018), Juliana Freire, Ioana Manolescu with four other male colleagues in (2020). The ACM Edgar F. Codd Innovations Award, also in databases initiated in 1992, has recognised Patricia Selinger (2002), Jennifer Widom (2007), Laura Haas (2015), and Anastasia Ailamaki (2019)—four in thirty years by 2022. The Internet recognises approximately thirty-six women who contributed to advances in computing; Wikipedia lists about sixty-four women’s contributions between 1842 and 2022 (in 180 years). Since the beginning of this science, outstanding contributions have been made by women in programming languages and programming (Fortran, Smalltalk C, C# Ruby). Many famous women programmers are clustered in video games, operating systems, software engineering and software evaluation. There is also an ambition to disseminate knowledge and active participation in computer education. These names do not include, for example, the recipients of the ACM SIGMOD Award Contributions or the ACM Edgar F. Codd Innovations Award. This situation demonstrates the dispersion when it comes to reconstructing history and accessing the memory of forgotten science when it comes to remembering the names of women scientists [43].

Missing datasets 3: Female contributions in CS organised by sub-discipline and geographic region. The discipline has different areas, including AI, DS, and much more, yet there is no database collecting significant contributions per discipline. Datasets are missing about the test of time and best paper awards concerning papers that have been highly cited during a ten-year interval. Indeed, major database conferences VLDB, SIGMOD/PODS, EDBT/ICDT, ICDE, and major data mining conferences like KDD have adopted this practice. Few databases collect information about papers with female partial or complete authorship, including, for example, the exhibition Women in Computing of the Science Museum in the UK (<https://www.sciencemuseum.org.uk/objects-and-stories/women-computing>, accessed on 1 June 2022). What happens at national conferences? What is their role in the local generation of significant knowledge? What is the part of female contribution to this knowledge production?

Missing datasets 4: Contributions in CS of female scientists of the global south.

We remark that these lists do not include women scientists working in the global south (i.e., Latin America and the Caribbean, Asia Pacific, Africa, and the Middle East). Who are the Latin American, Caribbean, Asian, and African women who have contributed to advances in computing science in different fields? What have been their contributions? How have their contributions improved their regions' development and knowledge? The first action by the United Nations recognised female scientists in Latin America and the Caribbean today. There is considerable work to collect data about female CS and STEM contributions.

2.4. Discussion

As a result of the identified datasets and observations done, we can derive answers to our research questions:

[RQ1] Do missing data concerning women's STEM research participation prevent a comprehensive view of their contribution across history?

Through the conclusions discussed in the three sources that we have analysed, we observed that women's history in STEM is partial and mainly located in the global north. We also observed that data are not organised according to disciplines and subdisciplines. This situation is particularly true in "young" sciences such as computer science and its subdisciplines. Regarding data science, if contributions to statistics and numerical methods should also be considered, it is true that very few and sparse data concern women, while women have been there contributing to mathematics and other sciences for ages. Data is sparse in that there are few or almost no datasets devoted to the topic, collected and built according to scientific methodologies. Any theory about women's contribution and absence in science history is anecdotal with these missing datasets.

[RQ2] Is it possible to design viable data science experiments that can estimate the female workforce in STEM and answer the question, "where are women in STEM"?

Data about the female workforce in STEM is provided in big grain, and the presence of women in different disciplines indeed changes a lot. The study presented in [44] shows, for instance, that the female authors in different computer science disciplines are not evenly distributed; areas like software engineering report more female publications co-authorships than in human performance architectures, for instance. With current datasets, even combining different datasets, such as awards and census of women scientists in universities, it is impossible to provide realistic cartography of women in the various STEM disciplines and subdisciplines. Besides, the location aspect remains since fewer datasets are available regarding the female workforce in the global south, even if they have widely contributed to STEM evolution. It is not realistic to design data science workflows for modelling women in STEM. Using, for example, clustering methods nor correlating variables that determine the choice of women choosing STEM careers and the evolution of their careers can find behavioural patterns that can explain their presence and absence in STEM. We should devote efforts to building datasets about women in STEM that are (i) intersectional, (ii) geographically representative, (iii) organised by discipline and subdiscipline, and (iv) show their experience in academia and industry.

3. The Mexican Case

Our vision is to address the study of missing datasets for studying the gender gap in STEM focuses on specific countries, considering that the context is essential to understanding the issue. Thus, we chose the Mexican case as a case study. The methodology adopted was to look for data about visible scientists in CS. The case study includes scientists recognised by the National System of Researchers (SNI), the Mexican Academy of Sciences, female scientists awarded by the National Award of Sciences, the history of Mexican women in STEM, and female professional situation and perspectives in STEM. We use these references because they provide an integrated national view of the Mexican

scientific community. A complete list of the acronyms used in this section is given in Appendix B.

3.1. Mexican National System of Researchers

According to their scientific production, the SNI classifies Mexican scientists with PhDs into five levels (candidate being the lowest one, and emeritus being the highest one). The evaluation focuses mainly on publications appearing in the ISI Thompson List and the impact factor of these publications. For the candidate level, applicants (less than 40 years old) must have published three journal papers recognized in the JCR or the list of journals of the CONACyT. To apply to level I, the scientist must have published 5 JCR papers, three of them in the last three years before the application. For level II, scientists must have published 15 JCR papers, 5 or 6 in the previous 5 or 6 years before the application. The scientist must also have advised graduate students. Finally, to apply to level 3, scientists must have published 15 to 30 JCR journal papers, and 8–9 papers must have been published during the last three years before the application. The awards can be renewed, and for SNI 3, emeritus scientists can have lifetime fellowship. Awarded scientists at SNI are granted fellowships that range between 375 USD– and 1750 USD.

Beyond the research are domains that cover a large spectrum of disciplines, organised in seven areas: I. mathematics, physics and earth sciences; II. biology and chemistry; III. medicine and health sciences; IV. humanities and behavioural sciences; V. social sciences; VI. biotechnology and agricultural sciences; and VII. engineering. What attracts attention is the distribution of the population across the SNI classification levels between men and women. The numbers show that 34% of Mexican scientists are women (!). What happened to all those women doing graduate studies, more significant in number than men? From this 34%, the majority of the female scientists are classified as grade I. This grade corresponds to people with at least 3 or 4 years of experience. Next, 15% of females are classified as grade II and 5% as grade III. These statistics show that it seems complicated for women to consolidate their careers and access recognised senior positions independently of the discipline. To gender inequality, we could add the age dimension, as Paloma Alcalá stated (Paloma Alcalá, *seminario “El sexo de la Ciencia”* held in the Faculty of Philosophy and Education Sciences of the Universidad del País Vasco, San Sebastián, 1 y 2 March 2000) that when women achieve the highest positions in the organisation, they have invested 16 or 20 years more than men. STEM, as other profession possibilities for women, are fields where women make less money and advance through the ranks more slowly [45]. Women’s absence at the top provides arguments to understand women’s underrepresentation in STEM.

During the new presidential six-year term (2018–2024), the SNI has made an effort in including women in evaluation committees. Recent raw data has been exported recently on the official governmental site; our study can be eventually completed with this new release. The gender gap is still wide, particularly in the highest grades of the SNI.

In 2014, the Mexican Council of Science and Technology (CONACyT) published the distribution of scientists doing research in the Mexican system recognised by the SNI. Only 15% of researchers in Mexico were women, with the criterion of having obtained the distinction of emeritus granted by the SNI. Later, according to official report of the CONACyT in 2020, there were 33,165 researchers recognised by the SNI, of which 8727 were candidates (26.31%), 17,091 level I (51.53%), 4793 level II (14.36%) and 2584 level III and emeritus (7.79%). In 2021, the percentage of women in the SNI achieved 38.2%. In 2022, the SNI granted a group of senior scientists with the emeritous grade: 183 scientists in all disciplines; of this number, 38 correspond to female researchers, the highest number of women who have been awarded this distinction. In 2022, there are 102 female researchers with emeritus status in the SNI out of a total of 462.

Let us analyse people doing research in computer science recognised in the SNI. We find eighty-two women researchers identified out of four hundred and thirty-seven recognised between candidate and three consecutive grades. There is only one female

scientist in grade three, Elba Patricia Melin Olmeda of the Instituto Tecnológico de Tijuana, a specialist in artificial intelligence, compared to thirteen male researchers. There are two female researchers in grade II, against thirty-eight male researchers with the same grade. The rest are distributed among forty female researchers at grade one against two hundred and seventeen male researchers. There are thirty-eight female researchers with grade “candidate” against ninety-seven male researchers. Regardless of observing an evident and pronounced gender disparity, sociological and economic studies are still needed to explain this phenomenon in the Mexican context.

The CONACyT and the Mexican government implemented inclusion policies seeking gender balance, promoting it in evaluation committees with the controversial idea that with more women in committees, they would encourage gender balance. In 2016, the SNI decision-making commissions, composed of 14 members each, were published on the CONACyT Web Site (<http://conacyt.gob.mx/index.php/convocatorias-conacyt/informacion-importante-sni/miembros-de-comisiones-dictaminadoras/12183-integrantes-de-las-comisiones-dictaminadoras-2016/file> (accessed on 20 August 2016)). The gender proportion was as follows:

- Physics, mathematics and earth sciences: 3 women:11 men;
- Biology, chemistry and life sciences: 5 women:9 men;
- Medicine and health sciences: 7 women:7 men;
- Humanities and behaviour sciences: 7 women:7 men;
- Social sciences: 3 women:11 men;
- Biotechnology and agricultural sciences: 3 women:11 men;
- Engineering: 1 woman:13 men;
- Technology: 2 women:12 men (female president).

In 2021, the distributions were as follows. Note that the classification of disciplines evolved, more women have the presidency of the commission, and their participation is higher in areas where they are underrepresented:

- Physics, mathematics and earth sciences: 3 women:11 men (male president);
- Biology, chemistry and life sciences: 8 women:6 men (male president);
- Medicine and health sciences: 11 women:3 men (female president);
- Behaviour sciences and education: 9 women:5 men (female president);
- Humanities: 9 women:5 men (female president);
- Social sciences: 10 women:4 men (female president);
- Agriculture, agriculture, forestry and ecosystem sciences : 5 women:9 men (female president);
- Engineering and technological development: 5 women:9 men (male president);
- Interdisciplinary: 5 women:9 men (male president).

Figure 2 shows a comparison between the number of female and male members of the evaluation commissions. The difference is still significant even if it lowered between 2016 and 2022. Note that the CONACyT merged engineering and technology and created a multidisciplinary area. This reorganisation of disciplines might be why the number of women in engineering and technology, for example, increased.

Missing datasets about the female scientific force in STEM at SNI: The Mexican government has adopted an open data initiative that includes the SNI. This initiative gives access to the SNI community, and it is possible to perform simple statistics about women in different disciplines. However, no contextual data can make it possible to perform more profound studies about the professional career vs. the evolution of SNI awarded grades. Are women working in universities located in urban areas? Are women working in institutions located close to the capital more likely to be awarded by the SNI? What kind of contact and working network do SNI-awarded women have? Does this network play a role in their evolution in the SNI grades? There are also missing datasets about the situations in which scientists, particularly women, lose the SNI grant for some time and then are granted again when their scientific production takes off. Additionally, datasets

about the metrics used for evaluation are missing. They could be helpful to observe how their values evolve according to gender and SNI level of scientists. This data could provide a more representative understanding of the “productivity” of scientists and its evolution correlated with the development of their career (e.g., administrative positions, classification in their institutions, projects coordination, student advising).

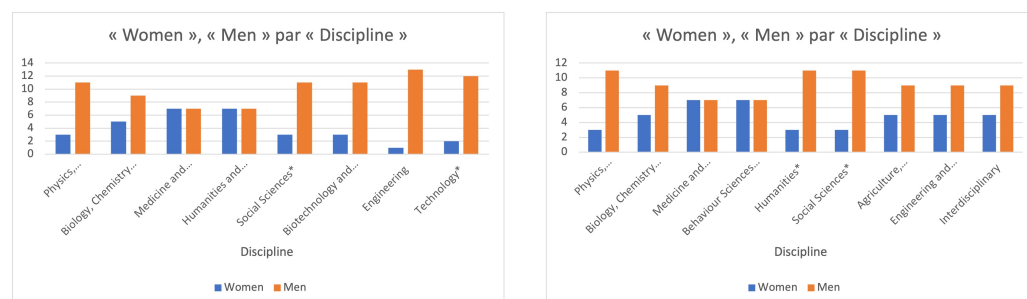


Figure 2. Comparison between the composition of commission members by gender and discipline between 2016 and 2020. Commissions with a * have female presidents. The complete names of the items with “...” are given in the text.

3.2. Mexican Academy of Sciences

Another reference showing the unbalanced number of men and women in STEM is the current membership numbers in the Mexican Academy of Sciences (AMC). The AMC is a non-profit organization comprised of distinguished Mexican scientists attached to various institutions in the country and several eminent foreign colleagues, including different Nobel Prize winners. According to the Web site of the AMC, by March 2022, there were 1376 members in “Exact Sciences” (astronomy, physics, engineering, chemistry and geosciences)—1158 men and 218 women. The discipline with fewer women was engineering, with 43 women and 318 men (see Figure 3).

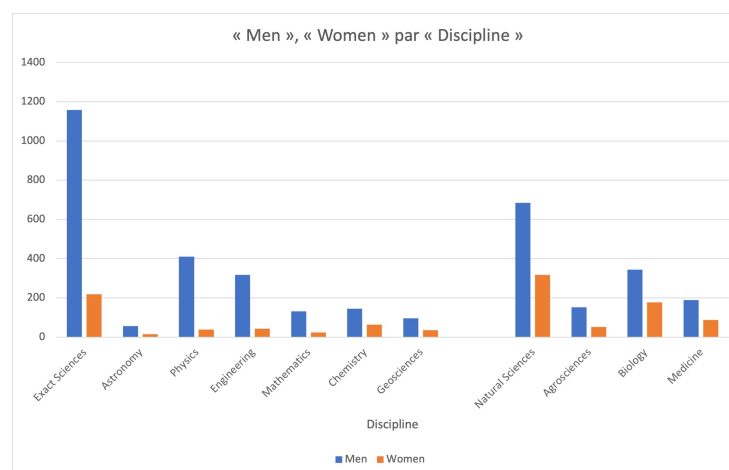


Figure 3. Members of the AMC by gender and discipline.

These numbers show that the AMC is even progressive considering that women’s access to university was granted recently. For example, in Europe, between 1860–1900, Cambridge accepted women without restrictions until 1947, and the National Academies of Sciences opened spaces for women later. Marjory Stephenson and Kathleen Lonsdale were the first scientists to join the Royal Society in 1945, an institution with 300 years of tradition. Yvonne Choquet-Bruhat was the first scientist accepted in the Academie de Sciences in France in 1979. This institution was founded in 1666. The first Spanish women received in the Royal Academy of Pharmacy and the Royal Academy of Sciences, Physics, and Natural Sciences were María Cascales (1987) and Margarita Salas (1988).

The AMC has already had its first female president, Rosaura Ruíz, who actively promoted the Science L'Oréal-UNESCO-AMC Award. She also encouraged the creation of fellowship programs for women studying Humanities and Social Sciences. She was awarded the National Award of Sciences and Arts in Technology and Design in 2009. In 2007, Esther Orozco was elected director of the Institute of Science and Technology of Mexico City, and she was awarded the L'Oréal UNESCO award in 2006. In 2009 Yoloxóchitl Bustamante was named the first female president of the Mexican Institute of Technology (Instituto Politécnico Nacional), 73 years after being founded. Some Mexican female scientists are influential and have occupied important positions in national and international institutions, including, for example, Ana María Cetto in the International Agency of Atomic Energy. She is a senior scientist at the Institute of Physics of the National Autonomous University of Mexico (UNAM). She has been head of the Faculty of Sciences and president of the Pugwash Conferences, a committee member that received the Nobel Peace Prize in 1995. In 1991, Ana María Cetto was president of the Organization for Women in Science for the Developing World (OWSD) for the Caribbean and Latin American Region. Mayra de la Torre was elected vice-president of the same institution. She collaborates in Science and Technology and the renewal of the program for Biotechnology/Biosecurity for the Americas. She has been awarded the National Award for Sciences and Arts, the Award of Sciences in the Third World, the Manuel Noriega Morales Award of the OEA and the CIBA GEIGY award in Innovation, Technology and Ecology.

Missing datasets about females in the AMC. Detailed datasets are missing regarding the contributions of women in the AMC. The open question is similar to data about women at the SNI: their connection networks, possible mentors, and mentoring roles. Which Mexican female scientists did they support to be accepted as members of the AMC?

3.3. Mexican National Award of Sciences

Let us take the Mexican National Award of Sciences as another possible data provider that shares little data about awardees. This award started being granted in 1945, and it recognises individuals or groups working in the following areas:

- Linguistics and literature beaux arts;
- History, social sciences and philosophy;
- Physics and mathematics, and natural sciences;
- Technology and design;
- Art and popular traditions.

Female scientists started to be awarded in 1979 with Guillermina Bravo, the first woman awarded for her work in Beaux-Arts. In 1996, María Luisa Ortega Delgado was awarded the National Award for Technology and Design with Adolfo Guzmán Arenas. In 2008, María de los Ángeles Valdés Ramírez was awarded (biology), and in 2009, Blanca Elena Jiménez Cisneros was awarded (environmental engineering) with José Luis Leyva Montiel. Thus, only one woman in CS has been awarded in the whole history of the awards program. Gender inequality is evident, and importance is given to gender issues in the country, considering that women acquired the right to vote in 1945 and that sexual liberation happened in the 60s.

The situation is even more precarious when making decisions and defining plans, policies and programs in committees. Indeed, few women participated in the structure of the Scientific and Technological Consultative Forum that coordinated the development of the National Science and Technology 2006–2012 (María Valdés, *La mujer mexicana en la ciencia*, https://www.cronica.com.mx/notas-la_mujer_mexicana_en_la_ciencia-946726-2016.html, accessed on 1 June 2022).

Missing datasets about female contributions in STEM and its implications. We have highlighted the lack of data about female contributions in STEM. The lack of data leads to the underrepresentation of women in visible databases that can help choose them to be granted awards. Their role is not highlighted in scientific teams and labs, project

coordination, keynote speeches, or innovation authorships. Therefore, they disappear from the scientific committees' memory, and thus, they are rarely granted prestigious awards.

3.4. History of Mexican Women in STEM

The history of Mexican women in science is still to be written and disseminated as part of the education of young Mexicans. Women have contributed to scientific production since ancient times. An example of one of the first Mexican women to attend University and obtain a diploma is Matilde Montoya (1859–1939). President Porfirio Díaz granted her permission to perform her professional examination, becoming the first Mexican Surgeon. Later, Helia Bravo (1901–2001) was the first Biologist in Mexico to develop scientific work on cactus. She published more than 160 papers and 3 books about these plants and the arid Mexican ecosystems. She described 57 species, 2 genders and 8 species that bear her name. Paris Pishmish (1911–1999) was an astronomer. She proposed a theory for explaining the origin and the spiral structure of galaxies. She discovered three open stellar cumuli, now named by her name. She also found that stellar galactic associations move away from the galaxy's centre. Then, Luz María del Castillo Fregoso, chemist and pioneer in biotechnology and zymology in Mexico, was internationally recognized for her work on physical chemistry, particularly on enzymatic reactions. She was the first woman to obtain the Sciences Award of the AMC, and she was a member of the SNI grade III from her first application.

Missing datasets about Mexican women in Science. The work of reconstructing the history of computing in Mexico from a gender perspective remains to be done. Naming all-female forgotten science contributors to STEM is important because this can let emerge forgotten activities and contributions that would be erased from the collective memory. It can make it possible to understand the conditions in which STEM fields produce results and capital, the real opportunities for professional careers, what people (men and women) should expect and how they can become agents of change.

3.5. Female Professional Situation and Perspectives in STEM

Women's representation varies by discipline and professorship status in the academic labour force. The majority of full-time female faculty in STEM disciplines are significantly low in computer and information sciences, maths, physical sciences, and engineering. In the life sciences, an area in which many people assume that women have achieved parity, women made up only one-third of faculties. The professional expectations do not seem to be very promising for women if we analyse the numbers reflecting female professionals accessing decision-making positions, including, for example, the number of female deans in engineering schools in universities and the number of university presidents in technology institutes globally. Of course, these numbers change among countries depending on culture, race, political trend and prominence of the institutions.

Studies show how women leaders in the US have a more significant impact on a company's bottom line (<https://www.fastcompany.com/90733328/the-secret-to-women-s-leadership-that-can-drive-such-a-positive-impact>, accessed on 1 June 2022). However, only an estimated 15% of C-suite executives and 51% of managers are women. Additionally, women make up only 30% of full professors and 26% of college deans in US universities (<https://www.catalyst.org/research/women-in-academia/>, accessed on 1 June 2022). However, in STEM, gender inequality is about leadership and inequality in terms of critical mass. It is impossible to choose female leaders where few women have the professional and/or scientific consolidation to access such positions. Referring to our use case, in Mexico, according to the Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES (<http://www.anui.es.mx>), accessed on 1 June 2022) between 2014–2015, 1 million 842 thousand 978 women and 1 million 876 thousand 17 men did studies in STEM fields. Thus, more or less, the numbers are balanced. Moreover, more women (167,967 women) than men do graduate studies in STEM (146,030 men). Therefore, what happens in professional life?

According to “Attitudes and experiences of engineering alumni” by [46–48] workplace environment, bias, and family responsibilities all play a role. Many women appear to encounter a series of challenges at mid-career that contribute to their leaving careers in STEM industries. Women cited feelings in studies of isolation, an unsupportive work environment, extreme work schedules, and unclear rules about advancement and success as significant factors in their decision to leave. Departmental culture includes the expectations, assumptions, and values that guide the actions of professors, staff, and students. Individuals may or may not be aware of the influence of departmental culture as they design and teach classes, advise students, organise activities, and take classes. For example, people tend to view women in “masculine” fields, such as most STEM fields, as either competent or likeable but not both, according to Madeline Heilman, an organizational psychologist at New York University [49,50]. Although being both competent and well-liked are essential for advancement in the workplace, this balance may be more difficult for women than men to achieve in science and engineering fields. STEM fields are perceived as male, including fields like chemistry and maths, where almost one-half of degrees awarded now go to women. Heilman’s research shows how, in the absence of clear performance information, individuals view women in male-type occupations as less competent than men [49,50]. There is a need to understand how people are evaluated in professional settings and understand which features in organisations give men more success opportunities. The first false assumption is that superior intelligence is required to address STEM fields. More than intelligence that can be numerically measured, the prejudice is that STEM calls for talent and that there is only one single talent that ensures success in such fields. The discrimination further argues about the possibility of numerically measuring such aptitude. The worst is the idea that such talent is a fixed gift, that talent is not malleable. Only those people with the highest standardised test scores and the most confidence will step into STEM fields and develop a career. Many believe that all these prejudices reduce the number of people deciding to make a career in STEM fields and even break the glass ceiling. The good news is that this talent is not a unitary thing. It is multidimensional and difficult to quantify and measure; many different skills are critical to step into STEM fields; talent can be developed and enhanced by education, encouragement, self-confidence, and hard work [45].

Few institutions promote reflection about policies for addressing gender issues in their organizations. In Mexico, there are three representative centres: the Centro de Investigaciones y Estudios de Género (CIEG) and the *Unidad Politécnica de Gestión con perspectiva de Género* (Polytechnic Unit for Management with a Gender Perspective) (<http://www.genero.ipn.mx/>, accessed on 1 June 2022) at the Universidad Autónoma de México and the Instituto Politécnico Nacional; and the Project on Gender Issues of the Humanities and Social Sciences School of the Monterrey Institute of Technology (Tecnológico de Monterrey). The CIEG has a unique chair on Gender in Science, Technology and Innovation, in addition to the Mexican Network of Science, Technology and Gender.

Missing datasets about working conditions of women in STEM. Women’s labour conditions when they develop a professional career in STEM are taboo both in academia and industry. Few or no data are collected by institutions, industry and independent bodies about the labour conditions: working hours, positions, evolution, environment, stress, entrepreneurship, glass ceiling, etc. How can the gender gap be understood and reduced with little input to drive studies and design policies? Data about these issues are disseminated in online forums (e.g., Quora, Medium, Glassdoor), blogs and social networks. They must be collected and integrated into datasets that can drive seminal studies.

3.6. Discussion

After applying a grounded study strategy for approaching datasets about women in STEM in general and then focusing on computer science, we observed that datasets are missing. The history of female contribution in STEM is often silent about women. We concluded that currently available data make it questionable to apply DM, ML and AI methods for picturing women’s history in STEM and discovering patterns to understand

their apparent absence. With these observations, we approximated conjectural answers for research questions RQ1 and RQ2. This first study could not provide elements to answer RQ3 because the studied datasets do not provide variables or perspectives about the conditions in which reported contributions were produced, only scientists' year, institution, discipline, and administrative nationality (nationality of birth and age). RQ3 questions about intersectionality, and therefore we chose a concrete case, that of Mexico, for which we observed the contribution statistics that show representative scientific production and excellence in the country. We enumerated the criteria used to evaluate scientific contribution in the National System of Researchers (SNI). The question was whether these criteria could provide elements to build an intersectional perspective and a quantitative study to approach an intersectional gender gap index.

[RQ3] Do missing intersectional perspectives prevent performing representative quantitative analysis about women's activity and contribution in STEM?

Existing datasets in Mexico providing data about women in STEM are sparse, partial and reduced to a tiny group of scientists (both female and male). For example, scientists working in private universities were excluded from the possibility of obtaining a fellowship related to SNI recognition. This exclusion hides an essential group of scientists from statistics. Universities, polytechnic centres, and other institutions that develop science provide different facilities to scientists to do their research. This difference depends on their location, type, and the disciplines they support. Female scientists and their conditions almost disappear in this complex context. Building intersectional datasets combining all these elements is a project to run. In a country such as Mexico, these socio-economic and political factors produce clusters of scientific production with different characteristics and conditions. Being a female scientist at the CINVESTAV (the most important research centre in Mexico) in Mexico City is entirely different than being a scientist in Huajuapán de León in Oaxaca (the poorest state in Mexico). The studies about women must address these aspects that call for thorough data collection strategies (by region and by the institution) and then running analytics at a small scale before scaling to the whole country. To answer RQ3, it is true that without intersectional datasets built with methodological data collection strategies, women will continue to be hidden and difficult to locate in the scientific cartography.

4. Conclusions

Gender gap issues in STEM are a matter of money and power and, more precisely, about capital in the sense of the theory of the French sociologist, Pierre Bourdieu (who defines three types of capital: education, material and cultural. Pierre Bourdieu proposes a complex machinery of interrelationships of these three types of capital that impose power and influence on each other and organise activities and social markets). Indeed, gender gap issues are all about an unbalanced problem among the different capitals activating STEM, particularly in promising disciplines like the AI and DS markets. These issues have been animated by bias associated with the roles and capacities related to men and women. Society believes and promotes that men are more apt to address STEM disciplines than women. However, the gender divide in STEM is more complex because of socio-cultural and racial factors. The official social order has viciously controlled access to specific knowledge disciplines.

The absence of women in STEM denotes the possibility of significant losses of brains with potential. If the gender perspective is not relevant or a priority from a human point of view, perhaps it can acquire value when it translates into millions of dollars lost by not valuing and encouraging the scientific contribution of a part of its population. In any case, this lack of balance in the STEM professional market attracts the attention of different actors willing to activate the global economy. For example, AI and DS are widely regarded as critical to the national economies in many "developed" countries. Concern about America's ability to be competitive in the global economy has led to several calls to action to strengthen the pipeline into these fields [51–53]. Thus, this unfair situation in countries of the global south like Mexico creates an economic loss. Countries where

inequality is more critical compete with less capital and thus with fewer possibilities to obtain benefits.

There is still a lot to be done, mainly because the capital conditions in which women as minorities enter the STEM professional and scientific market are very different from those in which men do. This issue is also related to education in STEM, the way young students perceive it and how they decide to step into STEM studies and careers depending on whether they are men or women. Many departments and universities worldwide develop studies about the aspects that move female students to inscribe in STEM majors and the conditions of their permanence from bachelor to graduate studies eventually. This collected data must be integrated and shared, but it has the merit of existing. Postcolonialism and decolonialism have already explained that minorities are never the same and cannot be studied and analysed as a whole. Racial aspects must be considered as well. In Mexico, indigenous women living in small villages cannot aspire to the same opportunities as women with more European racial heritage living in a big city. Nevertheless, the participation of women in STEM is still considered a special event that should have special treatment.

Data-driven studies are crucial to understanding the gender gap from an intersectional point of view, considering the particular conditions of the global north and south countries. This paper has shown that data are the fuel that can provide perspectives about the conditions in which women participate in the STEM market. The paper has also demonstrated that datasets are cruelly missing and that the challenge of collecting representative data is significant. Data-driven studies also need algorithms that can exploit them. These algorithms must be designed with care so that they do not bias observations and conclusions about the gender gap and do not further marginalise and discriminate against women in STEM. We need to open the door and allow and encourage women to contribute to the development of STEM, and this perspective depends on data and fair algorithms. Here is the view that should drive gender studies. Our current work addressed these issues in the context of the JOWDISAI and SINFONIA projects that are willing to create intersectional data collections about the women labour force in AI and DS in academia and industry in France. Other initiatives are seeking to reason about datasets, their content and collection conditions from decolonisation and feminist perspective, for example, the movement Tierra Común (<https://www.tierracomun.net>, accessed on 1 June 2022), and the A+ Alliance (<https://feministai.pubpub.org>, accessed on 1 June 2022). Our future work will include data collection strategies and algorithms to propose an intersectional analysis of the gender gap in STEM.

Funding: This research was partially funded by the projects JOWDISAI program AAP of the CNRS, France and SINFONIA of the interlaboratory program of the Institute of Gender, France.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

The following list shows contributions in CS authored by leading women. This list is complementary to the list of women granted with prestigious awards. The list shows the controversy about the relation between awards and contributions to knowledge with a gender perspective.

1. In 1843, Ada Lovelace (1815–1852) became the first programmer by designing the first algorithm and explaining how it would work in Babbage’s non-existent analytical engine.
2. In 1942, Hedy Lamarr invented the “frequency-hopping” technology that enabled the invention of wireless signals such as Wi-Fi and Bluetooth.

3. In 1945–46, Jean Bartik (1924–2011) and five women developed and coded the fundamentals of software programming working on the ENIAC. In spite of their decisive work, these women programmers were not invited to the dinner celebrating the construction of the ENIAC. These are their names: Frances Spence (1922–2012), Marlyn Meltzer (1922–2008), Kathleen Antonelli (1921–2006), Betty Holberton (1917–2001), Ruth Teitelbaum (1924–1986), and Gloria Gordon Bolotsky (1921–2009).
4. In 1951, Admiral Grace Hooper (1906–1992) created the first compiler for the programming language she proposed, COBOL.
5. Mary Allen Wikes (1937–), computer programmer and designer. Her best known work is on the LINC computer, now recognised as the first personal computer.
6. Lois Haibt (1934–) is known as one of the ten people who formed the IBM group that developed FORTRAN, the first successful high-level programming language.
7. Radia Perlman (1951–) is famous for inventing the spanning-tree protocol fundamental to network operation while working at Digital Equipment Corporation.
8. Elsie Shutt founded Computation Incorporated in 1957.
9. Danese Cooper (1959–) is a programmer, computer scientist, and activist in favour of open software.
10. Rebecca Heineman (1963–) is a video game programmer, veteran of the video game industry, founding member of Interplay Productions, Logicware, and Contraband Entertainment.
11. Jamie Fenton (1964–) is a programmer of the famous 1981 arcade hit game Gorf. In 1978, she created an early example of glitch art entitled Digital TV Dinner.
12. Corrinne Yu (1979–) is game programmer.
13. Audrey Tang (1981–) is a Taiwanese free software programmer described as one of the top ten Taiwanese computer programmers.
14. Lorinda Cherry, contributor to the Plan 9 Operating System, co-wrote with Mike Leks “Typing Documents on the UNIX System: Using the -ms and -mcs Macros with Troff” for the tenth edition of the Unix Handbook. She coded a non-dictionary-based spelling checker proposed by Bob Morris.
15. Xiaoyuan Tu is co-founder and lead scientist at AiLive Inc. (formerly iKuni Inc.) , a Silicon Valley startup working on artificial intelligence effects in computer entertainment. Dr. Tu has received awards for her technical contributions in the fields of computer animation, behavioural modelling and artificial life. In 1996, she received the prestigious ACM Doctoral Dissertation Award for her doctoral dissertation entitled “Artificial Animals for Computer Animation: Biomechanics, Locomotion, Perception, and Behavior”. Although other women have received awards for their doctoral dissertations, Dr. Tu is the first and only woman so far to receive first place.
16. Carla Meninsky was a video game designer and programmer during her early career at Atari VCS. Along with Carol Shaw, Meninsky was one of two engineers at Atari, Inc. who developed video game cartridges.
17. Leah Culver (1982–) co-founded the micro-blogging site Pownce which was bought by Six Apart in December 2008.
18. Amanda Wixted (1982–) is a game programmer and founder of Meteor Grove Software.
19. Dona Bailey is an American game programmer and educator who together with Ed Logg in 1981 created the game Centipede. Jade Raymond (1975–) is a Canadian video game executive and founder of Electronic Arts’ Motive Studios and Ubisoft Toronto.
20. Dina St. Johnston (1930–2007) was an English programmer credited with founding the first software house in the UK in 1959.
21. Linda Liukas (1986–) is a Finnish programmer, programming instructor and children’s book writer. In 2014, her children’s coding book “Hello Ruby” became the most financially supported book on Kickstarter.
22. Edith Windsor (1929–2017) was a chief technology officer at IBM.
23. Elaine Wayuker (1945–) is an ACM Fellow, IEEE Fellow, and AT&T Fellow at Bell Labs and a member of the National Academy of Engineering for her research in software metrics and evaluation.

24. Zoë Quinn (1987–) is a video game developer, programmer, writer and artist. She developed the interactive fiction *Depression Quest*.
25. Amy Briggs (1962–) is a video game programmer best known for creating “*Plundered Hearts*”, an interactive video game published by Infocom in 1987.
26. Molly Holzschlag (1963–) is an author, teacher and Open Web advocate. She has written and co-authored 35 books on Web design and open standards including “*The Zen of CSS Design: Visual Enlightenment for the Web*”.
27. Sandi Metz is a software engineer and author of “*Practical Object-Oriented Design in Ruby*”.
28. Ellen Ullman is a programmer and writer. She has written novels and articles for various publications including *Harper’s Magazine*, *Wired*, *New York Times* and *Salon*.
29. Coraline Ada Ehmke is a speaker, writer, open software advocate, technologist with over 20 years experience in Web application development. She is the creator of *Covenant Contributor*, the most popular open source code with 35,000 adoptions including JRuby, Swift, F#, GitLab, and Rails.
30. Brianna Wu (1980–) is a video game developer and programmer. She co-founded *Giant Spacekat*, an independent video game development studio with Amanda Warner in Boston, Massachusetts.

Appendix B

This section provides a list of acronyms with their definition.

- ACM—Association for Computing Machinery
- AMC—Mexican Academy of Sciences (Academia Mexicana de Ciencias)
- ANUIES—National Association of Universities and Institutions of Higher Education (Asociación Nacional de Universidades e Instituciones de Educación Superior)
- CIEG—Centro de Investigaciones y Estudios de Género
- CONACyT—National Council of Science and Technology (Consejo Nacional de Ciencia y Tecnología)
- EDBT—Extending Database Technology
- ICDE—International Conference on Data Engineering
- ICDT—International Conference on Database Theory
- IEEE—Institute of Electrical and Electronics Engineers
- ITESM—Monterrey Institute of Technology (Instituto Tecnológico de Monterrey)
- JCR—Journal Citation Reports
- KDD—Knowledge Discovery and Data Mining
- NCWIT—National Centre for Women & Information Technology
- OEA—Organisation of American States (Organización de los Estados Americanos)
- OWSD—Women in Science for the Developing World
- PODS—Principles of Database Systems
- SIGMOD—Special Interest Group on Management of Data
- SNI—National System of Researchers (Sistema Nacional de Investigadores)
- STEM—Science, Technology, Engineering and Mathematics
- UNAM—National Autonomous University of Mexico (Universidad Nacional Autónoma de México)
- VLDB—Very Large Data Base

References

1. Hill, C.; Corbett, C.; St Rose, A. *Why So Few? Women in Science, Technology, Engineering, and Mathematics*; American Association of University Women: Washington, DC, USA, 2010.
2. Keller, E.F.; Scharff-Goldhaber, G. *Reflections on Gender and Science*; American Association of Physics Teachers: College Park, MD, USA, 1987.
3. Lehming, R.F.; Alt, M.N.; Chen, X.; Hall, L.; Burton, L.; Burrelli, J.S.; Kannankutty, N.; Proudfoot, S.; Regets, M.C.; Boroush, M.; et al. *Science and Engineering Indicators 2010*; NSB 10-01; National Science Foundation: Arlington, VA, USA, 2010.

4. Eubanks, V. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*; St. Martin's Press: New York, NY, USA, 2018.
5. Faulkner, W. Doing gender in engineering workplace cultures. II. Gender in/authenticity and the in/visibility paradox. *Eng. Stud.* **2009**, *1*, 169–189. [[CrossRef](#)]
6. Young, E.; Wajcman, J.; Sprejer, L. *Where Are the Women? Mapping the Gender Job Gap in AI*; Policy Briefing: Full Report; The Alan Turing Institute: London, UK, 2021.
7. Berman, F.D.; Bourne, P.E. Let's make gender diversity in data science a priority right from the start. *PLoS Biol.* **2015**, *13*, e1002206. [[CrossRef](#)]
8. Alfrey, L.; Twine, F.W. Gender-fluid geek girls: Negotiating inequality regimes in the tech industry. *Gend. Soc.* **2017**, *31*, 28–50. [[CrossRef](#)]
9. Margolis, J.; Fisher, A. *Unlocking the Clubhouse: Women in Computing*; MIT Press: Cambridge, MA, USA, 2002.
10. Bulmahn, E. Women in science in Germany. *Science* **1999**, *286*, 2081. [[CrossRef](#)]
11. Costas, I. Women in science in Germany. *Sci. Context* **2002**, *15*, 557–576. [[CrossRef](#)]
12. Mello, P.A. *Qualitative Comparative Analysis: An Introduction to Research Design and Application*; Georgetown University Press: Washington, DC, USA, 2022.
13. Makarova, E.; Aeschlimann, B.; Herzog, W. The gender gap in STEM fields: The impact of the gender stereotype of math and science on secondary students' career aspirations. *Front. Educ. Front.* **2019**, *4*, 60. [[CrossRef](#)]
14. Fayyad, U.; Hamutcu, H. Toward Foundations for Data Science and Analytics: A Knowledge Framework for Professional Standards. 2020. Available online: <https://hdsr.mitpress.mit.edu/pub/6wx0qmk1/release/4?readingCollection=70ac5c46> (accessed on 1 June 2022).
15. Kemp, P.E.; Wong, B.; Berry, M.G. Female performance and participation in computer science: A national picture. *ACM Trans. Comput. Educ. (TOCE)* **2019**, *20*, 1–28. [[CrossRef](#)]
16. Maltese, A.V.; Cooper, C.S. STEM pathways: Do men and women differ in why they enter and exit? *AERA Open* **2017**, *3*, 2332858417727276. [[CrossRef](#)]
17. D'ignazio, C.; Klein, L.F. *Data Feminism*; MIT Press: Cambridge, MA, USA, 2020.
18. Hearn, J.; Louvrier, J. Theories of difference, diversity, and intersectionality. In *The Oxford Handbook of Diversity in Organizations*; Oxford University Press: Oxford, UK, 2015; p. 62.
19. Benjamin, R. Race after technology: Abolitionist tools for the new jim code. *Soc. Forces* **2019**, *98*, 1–3. [[CrossRef](#)]
20. Herring, C. Does diversity pay? Race, gender, and the business case for diversity. *Am. Sociol. Rev.* **2009**, *74*, 208–224. [[CrossRef](#)]
21. Collins, P.H. It's all in the family: Intersections of gender, race, and nation. *Hypatia* **1998**, *13*, 62–82. [[CrossRef](#)]
22. Bobbitt-Zeher, D. Gender discrimination at work: Connecting gender stereotypes, institutional policies, and gender composition of workplace. *Gend. Soc.* **2011**, *25*, 764–786. [[CrossRef](#)]
23. Wajcman, J.; Young, E.; Fitzmaurice, A. *The Digital Revolution: Implications for Gender Equality and Women's Rights 25 Years after Beijing*; United Nations: New York, NY, USA, 2020.
24. Myers West, S. *Discriminating Systems: Gender, Race and Power in Artificial Intelligence*; Georgia Institute of Technology: Atlanta, GA, USA, 2020.
25. Freire, A.; Porcaro, L.; Gómez, E. Measuring diversity of artificial intelligence conferences. In Proceedings of the Artificial Intelligence Diversity, Belonging, Equity, and Inclusion, Online, 21–31 January 2021; pp. 39–50.
26. Abbate, J. *Recoding Gender: Women's Changing Participation in Computing*; MIT Press: Cambridge, MA, USA, 2012.
27. Alegria, S. Escalator or step stool? Gendered labor and token processes in tech work. *Gend. Soc.* **2019**, *33*, 722–745. [[CrossRef](#)]
28. UNESCO. *Artificial Intelligence and Gender Equality: Key Findings of UNESCO'S Global Dialogue*; UNESCO: Paris, France, 2020.
29. Cardador, M.T.; Hill, P.L. Career paths in engineering firms: Gendered patterns and implications. *J. Career Assess.* **2018**, *26*, 95–110. [[CrossRef](#)]
30. Ensmenger, N.L. *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise*; MIT Press: Cambridge, MA, USA, 2012.
31. Foulds, J.R.; Islam, R.; Keya, K.N.; Pan, S. An intersectional definition of fairness. In Proceedings of the 2020 IEEE 36th International Conference on Data Engineering (ICDE), Dallas, TX, USA, 20–24 April 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1918–1921.
32. Simonite, T. AI Is the Future—But Where Are the Women. 2018. Available online: <https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance> (accessed on 1 June 2022).
33. Dobbin, F.; Kalev, A. Why diversity programs fail. *Harv. Bus. Rev.* **2016**, *94*, 14.
34. Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the Conference on Fairness, Accountability and Transparency, New York, NY, USA, 23–24 February 2018; pp. 77–91.
35. Wajcman, J. *Feminism Confronts Technology*; Penn State Press: University Park, PA, USA, 1991.
36. Wajcman, J. Feminist theories of technology. *Camb. J. Econ.* **2010**, *34*, 143–152. [[CrossRef](#)]
37. Allen, K.J. Hidden Figures: The American Dream and the Untold Story of the Black Women Mathematicians Who Helped Win the Space Race by Margot Lee Shetterly. *IEEE Ann. Hist. Comput.* **2017**, *39*, 70–71. [[CrossRef](#)]
38. Marino, K. The Glass Universe: How the Ladies of the Harvard Observatory Took the Measure of the Stars. *Hist. J. Mass.* **2018**, *46*, 172–177.

39. Light, J. Rise of the Rocket Girls: The Women Who Propelled Us, from Missiles to the Moon to Mars. *Nature* **2016**, *532*, 34–35. [[CrossRef](#)]
40. Light, J.S. When computers were women. *Technol. Cult.* **1999**, *40*, 455–483. [[CrossRef](#)]
41. Coello, C.A.C. *Breve Historia de la Computación y Sus Pioneros*; Fondo de Cultura Económica: México City, Mexico, 2003.
42. Copeland, B.J. The Modern History of Computing. 2000. Available online: <https://plato.stanford.edu/entries/computing-history/> (accessed on 1 June 2022).
43. Galpin, V. Women in computing around the world. *ACM SIGCSE Bull.* **2002**, *34*, 94–100. [[CrossRef](#)]
44. Vela, B.; Caverio, J.M.; Vargas-Solar, G.; Espinosa-Oviedo, J.A.; Cáceres, P. A Geo-Gender Study of Indexed Computer Science Research Publications. *arXiv* **2021**, arXiv:2105.00972.
45. Valian, V. *Why So Slow? The Advancement of Women*; MIT Press: Cambridge, MA, USA, 1999.
46. Hewlett, S.A.; Luce, C.B.; Servon, L.J.; Sherbin, L.; Shiller, P.; Sosnovich, E.; Sumberg, K. The Athena factor: Reversing the brain drain in science, engineering, and technology. *Harv. Bus. Rev. Res. Rep.* **2008**, *10094*, 1–100.
47. Frehill, L.; Javurek-Humig, A.; Jeser-Cannavale, C. Women in Engineering: A review of the 2005 literature. *Mag. Soc. Women Eng.* **2006**, *52*, 34–63.
48. Interactive, H. Attitudes and Experiences of Engineering Alumni, Prepared for the Society of Women Engineers. *unpublished*.
49. Heilman, M.E.; Wallen, A.S.; Fuchs, D.; Tamkins, M.M. Penalties for success: Reactions to women who succeed at male gender-typed tasks. *J. Appl. Psychol.* **2004**, *89*, 416. [[CrossRef](#)]
50. Heilman, M.E.; Okimoto, T.G. Why are women penalized for success at male tasks? The implied communality deficit. *J. Appl. Psychol.* **2007**, *92*, 81. [[CrossRef](#)]
51. McClure, C.R. *A Test of Leadership: Charting the Future of US Higher Education*; U.S. Department of Education Contract: Jessup, MD, USA, 2007.
52. Committee on Prospering in the Global Economy of the 21st Century. *Rising above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*; National Academies Press: Washington, DC, USA, 2007.
53. Ashby, C. Science, technology, engineering, and mathematics: Trends and the role of federal programs. In *Testimony before the Committee on Education and the Workforce, House of Representatives*; National Center on Educational Outcomes (NCEO): Minneapolis, MN, USA, 2006.