

Bridging the Divide: Gender, Diversity, and Inclusion Gaps in Data Science and Artificial Intelligence Across Academia and Industry in the majority and minority worlds

Genoveva Vargas-Solar^{1†}

genoveva.vargas-solar@cnrs.fr

^aCNRS, Univ Lyon, INSA Lyon, UCBL, LIRIS, UMR5205, F-69221, France

Abstract

As Artificial Intelligence (AI) and Data Science (DS) become pervasive, addressing gender disparities and diversity gaps in their workforce is urgent. These rapidly evolving fields have been further impacted by the COVID-19 pandemic, which disproportionately affected women and minorities, exposing deep-seated inequalities. Both academia and industry shape these disciplines, making mapping disparities across sectors, occupations, and skill levels essential. The dominance of men in AI and DS reinforces gender biases in machine learning systems, creating a feedback loop of inequality. This imbalance is a matter of social and economic justice and an ethical challenge, demanding value-driven diversity. Root causes include unequal access to education, disparities in academic programs, limited government investments, and underrepresented communities' perceptions of elite opportunities. This chapter examines the participation of women and minorities in the fields of AI and DS, with a focus on their representation in both industry and academia. Analysing the existing dynamics seeks to uncover the collective and individual impacts on the lives of women and minority groups within these fields. Additionally, the chapter aims to propose actionable strategies to promote equity, diversity, and inclusion (DEI), fostering a more representative and supportive environment for all.

1. Introduction

As promising domains in Information Technologies (IT), such as artificial intelligence (AI) and data science (DS), become increasingly ubiquitous, addressing gender disparities and diversity, equity and inclusion (DEI) gaps in the workforce of these disciplines is urgent and necessary. The rapid

evolution of AI and DS, coupled with the disproportionate impact of the COVID-19 pandemic on women and minorities, has exacerbated existing inequalities, further exposing systemic challenges [10, 23–25]. Industry and academia play an essential role in shaping these fields, making it critical to map how these disparities manifest across different sectors, occupations, and skill levels. This underscores the urgent need to critically examine these issues and implement strategies to mitigate disparities and foster a more inclusive workforce.

The persistent dominance of men in AI and DS, especially across various regions, contributes to a feedback loop that perpetuates gender biases within machine learning and AI systems. This unbalanced distribution of opportunities is fundamentally an issue of social and economic justice and an ethical challenge, highlighting the need for value-driven diversity. Understanding the root causes of this discriminatory landscape requires examining factors such as unequal access to education, disparities in the availability of undergraduate and graduate programs in these disciplines, variations in government investments in technology and education, and the perceptions underrepresented communities have about accessing elite opportunities.

This chapter examines the participation of women and other minorities in IT like DS and AI within industry and academia across majority and minority contexts to explore their individual and collective impacts on women’s lives. Accordingly, the chapter is organised as follows. Section 2 explores gender representation in AI and DS, comparing workforce participation across different countries, particularly within the European context, and analysing structural barriers such as mentorship access, funding, and cultural biases. It examines the historical biases embedded in AI and DS, the influence of societal perceptions on career choices, career progression obstacles, and the exacerbation of these challenges due to the COVID-19 pandemic. Section 3 highlights the importance of gathering diverse workforce data and proposing various collection methods, such as surveys and industry reports, while addressing ethical considerations, including privacy and bias mitigation. Sections 4 and 5 introduce and illustrate key workforce metrics such as turnover and attrition rates, comparing industry and academia and evaluating role-specific trends in managerial, technical, and research positions. It proposes solutions such as educational initiatives, workplace interventions like mentorship programs, inclusive hiring policies, and recommendations for industry-academia collaboration to foster greater diversity. Finally, section 6 summarises the paper’s key findings, suggests directions for future research, and calls for concerted efforts from stakeholders to reduce gender disparities in AI and DS, ultimately fostering a more inclusive workforce.

2. Women in economically promising IT domains

Given the near-ubiquitous presence of AI and DS in modern society, addressing gender disparities and gaps in DEI within these fields has become imperative. We focus on the French and European contexts to discuss the participation of women in economically promising IT domains.

France presents a unique case study as one of the leading European economies. The country has a rich technological ecosystem supported by government initiatives and an intense academic landscape. However, gender disparities persist, particularly in economically promising fields such as AI, DS, and other STEM-related industries. Despite governmental efforts to enhance gender diversity, women remain underrepresented in these fields, both in higher education and the workforce. The fields are particularly fast-moving both in industry and academia, so it is essential to comprehensively map how these gaps [4, 31] are manifest across different sectors, occupations, and skills. While women represent nearly half of university students in France, their participation in AI and DS remains significantly low. Reports indicate that only around 15% of AI specialists in France are women, reflecting a stark gender imbalance. Structural barriers, including cultural biases, lack of mentorship, and limited access to funding and networking opportunities, contribute to this disparity.

Compared to other European countries, France falls behind leaders such as Sweden and Finland, where gender diversity initiatives in STEM fields have seen relative success [23]. These countries have implemented gender quotas, funding for female-led tech startups, and dedicated programs to encourage women’s participation in IT-related disciplines. Conversely, countries like Germany and Italy also struggle with low female representation in IT, particularly in higher academia and research positions, mirroring the challenges observed in France [25]. The representation of women in AI-related fields is particularly low in Southern and Eastern Europe, with countries such as Greece and Poland reporting some of the widest gender gaps in IT employment [2]. However, women in France have been more likely to face job losses in tech-related fields due to the pre-existing gender gap in digital skills. The pandemic also intensified domestic and caregiving responsibilities, limiting women’s ability to engage in reskilling programs or advance their careers in high-tech industries.

This ubiquitous male dominance [13, 21, 28] in countries in different regions results in a feedback loop shaping gender bias in AI and machine learning systems [37, 38, 72]. AI algorithms are often trained on datasets that reflect historical gender biases, reinforcing and perpetuating systemic

inequities. The underrepresentation of women in AI development means critical perspectives are missing to challenge and correct these biases. Without intentional intervention, the gender gap in AI and DS will continue to grow, ultimately affecting the fairness and Inclusivity of technological advancements.

We believe that the phenomenon should be studied by country and then scaled to more global studies. In France, policies that promote gender-inclusive education, equitable hiring practices, and targeted mentorship programs are necessary to bridge the gap. Encouraging women’s participation in AI and DS is an ethical and economic imperative—diverse teams drive innovation and create more robust AI models that serve a broader spectrum of society. Addressing these challenges will ensure France remains competitive in the global AI landscape while fostering an inclusive and fair technological future.

2.1. AI and DS female workforce

The persistent absence of women in AI and DS fields is troubling. According to the report of the World Economic Forum in 2018, over three-quarters of professionals in these fields globally are male (78%); less than a quarter are women (22%) [12]¹. What about other underrepresented communities [7, 9]? How are they represented in the DS and AI workforces, and what are some of the opportunities offered by these promising areas taken by these communities? Of course, to acquire a full understanding of this phenomenon, it is necessary to treat the female community [35, 46, 62]. Other underrepresented communities [9, 45] are multifaceted and heterogeneous groups with a plurality of experiences, where gender intersects with multiple aspects of difference and disadvantage [22].

Different academic organisations like IEEE (Women in Engineering), ACM (women in ACM) in the US, the European Institute of Gender Equality in Europe, the CNRS (Centre National de la Recherche Scientifique) 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, Google², 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

¹World Economic Forum. Global Gender Gap Report 2018. Retrieved from: <https://www.weforum.org/reports/the-global-gender-gap-report-2018>.

²Google (2020). Google Diversity Annual Report 2020. Google.

Mathematics) community. The opportunities, economic share, and return on investment in different disciplines in STEM are not homogeneous. In the last decade, the role of data and the scientific and technical skills used to exploit it have created promising careers and economic spaces. Artificial Intelligence (AI) and Data Science (DS) have emerged as promising areas for developing careers with critical economic benefit perspectives. Nevertheless, professional career perspectives in these disciplines are not equal depending on gender ³ [68, 70] and other criteria [42], including race/ethnicity, socioeconomic level, the institution's reputation where people were educated, etc. For example, the House of Lords Select Committee on Artificial Intelligence in the United Kingdom 2018 advocated increasing gender and ethnic diversity among AI developers. In France, several companies like Renault and Engie, through the "Laboratoire de l'Egalité", 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 [23–25] noted that it is 'high time to reflect precisely on the interplay between AI and gender equality' [23]. French, European [7], and international organisations and agencies [1, 3, 6] perform studies for observing workforce shares in industry and sometimes in academia in a global perspective, often observing STEM [52]. Few fine grain studies have studied underrepresented communities' workforce evolution and gaps from the gender perspective in promising disciplines like AI and DS [11, 51, 63, 72]. A thorough understanding of the way the workforce accesses the AI and DS opportunities in industry [17, 18] and academia [36, 41] is essential for building fair and inclusive societies [48, 55], but also for making sure that countries obtain full benefits of the development of these areas to achieve better economic and social conditions and leading positions in the international arena through technology self-sufficiency.

³Contributes to Research and Innovation. Policy Review. H2020 Expert Group will update and expand 'Gendered Innovations/ Innovation through Gender'. Luxembourg: Publications Office of the European Union.

2.2. Challenges

Despite the economic and symbolic capital ⁴ investment seeking a fair distribution of AI and DS opportunities for women and underrepresented communities, the crystal ceiling must still be reached and broken [32]. Part of the explanation resides in data (!) [15, 30]. Indeed, many actors and studies agree that quality, disaggregated, intersectional data are still missing, and these data are essential to interrogate and tackle inequities in the AI and data science labour force [62]. As stated in the Alan Turing Institute study, "Where are women?" [72], 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 priorities in making data-centred decisions to determine where to invest and ensure that underrepresented groups are given equal opportunity.

We believe it is necessary to promote actions by data experts applying mathematical and machine learning techniques for building complete, high-quality, privacy-preserving data collections that can be used for performing studies that can drive conclusions about gender and DEI gaps in these disciplines.

To achieve a comprehensive and representative understanding of the dynamics shaping the status and perspectives of women in economically promising IT domains, it is essential to integrate both social and quantitative dimensions. This includes examining societal factors, such as patriarchal perceptions of gender roles, alongside measurable data. Collecting relevant data and establishing robust metrics to assess gender and qualification gaps within IT fields is crucial. These metrics should be systematically incorporated into curation and analytics pipelines, enabling ongoing monitoring and informed decision-making to address disparities and promote Inclusivity.

3. Towards a curated and intersectional data collection

Understanding the professional landscape of women in AI, DS, Blockchain, and other economically promising IT expertise areas requires a robust data curation approach. The process involves systematically harvesting, integrating, archiving, and curating diverse digital and non-digital data sources that

⁴Symbolic capital, according to Pierre Bourdieu, is the prestige, recognition, and legitimacy an individual or group holds within a social field. It is a misrecognized form of other types of capital (economic, cultural, or social), gaining value through collective acknowledgment and legitimization. Symbolic capital grants authority and influence without direct coercion, shaping power dynamics in areas like academia, politics, and the arts.

may not always be directly available. The primary goal is to construct comprehensive datasets that accurately represent the composition of the workforce, career progression, and opportunities while considering variables such as gender, ethnicity, socioeconomic status, and sexual orientation [49, 56]. We consider that there are two main research questions introduced by the challenge of curating datasets considering intersectional⁵ aspects:

- RQ1: How to build, curate and maintain an intersectional data collection integrating data from available sources about workforce in AI and DS respecting privacy and General Data Protection Regulation (GDPR)?
- RQ2: How to compute a quantitative profile of the gender and diversity and inclusion gap in AI and DS professional careers according to a methodology that can lead to the enlightening interpretation of results?

3.1. Existing Data

The evidence base about gender diversity in the AI and DS workforce is minimal [8]. The available data is fragmented, incomplete and inadequate for investigating the career trajectories of women and men in the fields [30]. Available datasets 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. Besides, available data about the global AI and DS workforce is often aggregated and rarely broken down by age, race, geography, (dis)ability, sexual orientation, socioeconomic status, and gender. As stated by [66, 67, 72]:

"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".

⁵Intersectionality is a sociological framework that provides a nuanced lens for analysing how overlapping social and political identities—such as gender, race, ethnicity, class, sexuality, religion, disability, age, caste, physical appearance, height, and weight—intersect to create distinct experiences of discrimination and privilege for individuals and groups. This approach highlights the complexity of systemic inequalities by emphasising how multiple dimensions of identity interact to shape unique lived realities <https://en.wikipedia.org/wiki/Intersectionality>.

Estimating the Number of women working in AI and DS across various platforms and regions is challenging due to limited publicly available data. However, existing studies and reports provide some insights:

- Women constitute approximately 22% of AI professionals worldwide. This underrepresentation is evident across various regions and platforms [69].
- In Germany, women make up about 20,3% of AI professionals, which is below the international average. This places Germany among the countries with lower female representation in AI within Europe [69].
- The United States leads with approximately 442,000 AI professionals, followed by India with 213,000, the United Kingdom with 69,000, and Germany with 58,000. While specific gender breakdowns for these countries are not provided, the global trend of underrepresentation suggests similar patterns [69].
- Platform-Specific Insights:
 - Research indicates gender biases in participation and knowledge sharing on Stack Overflow, a major platform for programmers. Women are underrepresented, which may affect their visibility and engagement in the community [57].
 - While specific numbers are scarce, studies suggest that women are underrepresented on GitHub, a leading platform for code sharing and collaboration. This underrepresentation can impact collaborative opportunities and project contributions [33].

Accurately quantifying the Number of women in AI and DS roles across platforms like LinkedIn, ORCID, Google Scholar, Kaggle, Stack Overflow, and GitHub is difficult due to several factors:

- Self-Reported Data: Many platforms rely on users to self-report gender, leading to incomplete datasets.
- Privacy Concerns: Users may choose not to disclose personal information, making it hard to gather comprehensive statistics.
- Dynamic Nature: The rapidly evolving tech landscape means data can quickly become outdated.

Although precise statistics remain challenging, existing research consistently highlights a pronounced underrepresentation of women in AI and DS across diverse regions and platforms. Addressing this disparity is critical to advancing DEI within the tech industry [31, 48, 55]. Understanding the extent of women’s participation in these fields, their economic contributions, and their share of the financial benefits is essential for creating a more equitable and inclusive workforce. This analysis sheds light on the current gaps and underscores the broader economic and social advantages of fostering gender parity in AI and DS.

3.2. Intersectionality and Inclusion in Workforce Data Collection

From a Social and Human Sciences (SHS) perspective, it is essential to define Intersectionality in the AI and DS workforce by incorporating DEI aspects. A feminist, inclusive perception of data must guide the selection and analysis of workforce trends [49, 50]. A crucial element of this approach involves defining inclusion/exclusion criteria for data sources to ensure an intersectional data collection that integrates various marginalised and underrepresented groups.

Harvesting such data presents additional challenges since official documents and open data from companies, governmental agencies, and academia may lack direct information on these variables. Thus, developing methodologies to extract implicit information and infer missing demographic insights from structured and unstructured data becomes critical [56].

From an IT perspective, data integration must balance access to private and sensitive information (e.g., gender, race, salary, sexual orientation, and socioeconomic status) and individual privacy protection. Strategies for ethically managing sensitive data must adhere to regulations such as the GDPR [14], ensuring that the collected datasets remain anonymised and comply with legal and ethical standards.

Furthermore, ensuring data reinforcement strategies—such as bias detection, missing data imputation, and real-time updates—requires semi-automated solutions that integrate governance policies with algorithmic Transparency and explainability. AI-driven data curation methods must also ensure traceability and accountability for the sources included in the dataset.

3.2.1. Principles of Intersectional and Inclusive Data Collection

Our data collection approach is guided by the following:

- **Intersectionality:** Recognising the overlapping influences of gender, race, ethnicity, disability, and socioeconomic background on career opportunities.

- **Inclusivity:** Ensuring all workforce demographics, including under-represented groups, are represented.
- **Transparency:** Providing transparent methodologies and ethical guidelines for data collection.
- **Privacy and Consent:** Ensuring compliance with data protection laws (e.g., GDPR) and obtaining informed consent from participants.

We propose a multifaceted approach to gather comprehensive workforce data:

1. *Surveys and Self-Reporting*

- Conduct voluntary, anonymous surveys covering demographic, educational, and career-related aspects.
- Use standardised questions across platforms to ensure consistency.
- Offer multilingual survey options to increase participation across diverse populations.

2. *Integration with Professional Platforms*

- Leverage data from LinkedIn, ORCID, Google Scholar, and industry recruitment platforms.
- Employ AI-driven parsing tools to analyse skill sets, job mobility, and pay disparities.

3. *Employer and Institutional Reporting*

- Collaborate with IT companies and academic institutions to obtain workforce diversity reports.
- Encourage organisations to report gender pay gap analyses, recruitment trends, and retention statistics.

4. *Data from Government and Public Records*

- Utilise labour market data from national statistics agencies.
- Compare workforce participation rates across different socioeconomic backgrounds.

Key Metrics for Workforce Analysis To identify disparities and trends, data collection should focus on:

- **Demographics:** Gender, ethnicity, disability status, and socioeconomic background.
- **Educational Background:** Degree levels, institutions attended, access to STEM programs.
- **Employment Trends:** Role types, promotion rates, wage gaps, attrition rates.
- **Skill Development:** Training participation, certification completion, mentorship access.
- **Workplace Environment:** Employee satisfaction, harassment reports, inclusion initiatives.

Ethical Considerations Ensuring ethical workforce data collection requires:

- **Informed Consent:** Participants must understand data use and opt-in voluntarily.
- **Anonymisation:** Personal identifiers should be removed to protect privacy.
- **Bias Mitigation:** Implement measures to ensure diverse representation in datasets.
- **Regulatory Compliance:** Align with legal frameworks such as GDPR, EEOC, and labour policies.

3.3. Curation for designing intersectional women careers on IT dataset

Our methodology involves collecting and analysing professional profiles and career trajectories from prominent AI and DS platforms, including LinkedIn, ORCID, Google Scholar, Kaggle, StackOverflow, and GitHub [64]. The primary goal is to establish an intersectional data framework that captures detailed insights into professional careers, such as positions, profiles, and responsibilities while tracking French professionals’ evolution and representation across these platforms. By integrating data from professional social networks (LinkedIn, ORCID, Google Scholar) and technical platforms (Kaggle, StackOverflow, GitHub, and MS Scientific Graph), we aim to provide a comprehensive understanding of workforce dynamics and identify patterns of participation and advancement in AI and DS fields. This approach will enable a deeper exploration of DEI within these domains.

Regarding intersectional variables, according to the data provided by the profiles of these platforms, it will be possible to consider the geographic location of professional activities, companies, studies, topics, gender, AI and DS topics/projects, and active career duration. In some cases, it will be possible to oppose the country(ies) where a person studied, worked and where they originated. This information can give a hint on race/ethnicity. Regarding studies, the person studying in private/public institutions with a specific reputation (according to, for instance, the Shanghai ranking) can give hints on possible sociocultural levels or membership to elite intellectual communities.

The curation of an intersectional dataset on women’s careers in IT is a critical step toward fostering DEI in AI and DS. By integrating ethical data collection practices and leveraging analytical tools, this initiative can provide valuable insights to contribute to a more inclusive technology landscape. Moving forward, continuous updates and interdisciplinary collaborations will be essential to ensure the dataset remains relevant and impactful.

3.4. Data curation pipeline

The curation pipeline and metadata schema provide a comprehensive structure to understand the representation of women in IT. The data curation process follows a structured pipeline comprising six key stages (see Figure 1):

1. Data acquisition: To construct a comprehensive dataset, raw career data is gathered from various platforms, such as LinkedIn, ORCID, Google Scholar, Kaggle, Stack Overflow, and GitHub, to ensure a broad representation of professional and academic activities. Additionally, publicly accessible records—including research publications, patents, and open-source contributions—are systematically collected

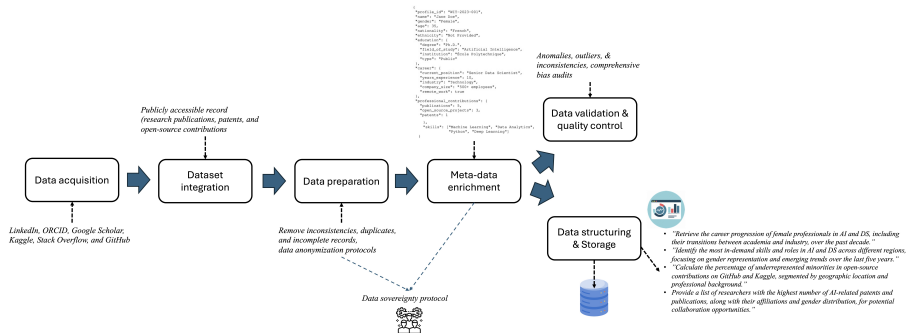


Figure 1: Data curation pipeline

and integrated into the dataset. This multi-source approach ensures a robust and holistic foundation for analysis, enabling a thorough examination of career trajectories, contributions, and representation across the fields of AI and DS.

2. **Data Preprocessing:** The collected data is standardised into a unified format to ensure consistency and compatibility across all platforms. This process involves cleaning and normalising entries to remove inconsistencies, duplicates, and incomplete records, thereby improving the overall quality and reliability of the dataset. Additionally, rigorous data anonymisation protocols are applied to protect sensitive information, ensuring full compliance with privacy regulations and ethical standards. This step is essential for creating a clean, secure, usable dataset ready for in-depth analysis.
3. **Metadata Enrichment:** Enhance raw data by adding structured metadata fields, including:
 - **Personal Attributes:** Gender, age, nationality, ethnicity (where available and ethically collected).
 - **Education and Training:** Degree levels, institutions attended, field of study, type of institution (public/private).
 - **Career Progression:** Job titles, employment transitions, industry versus academia affiliations.
 - **Geographic Mobility:** Countries of study, employment, and professional affiliations.
 - **Professional Contributions:** Research publications, patents, repositories, collaborations, leadership roles.
 - **Skills and Expertise:** Areas of specialisation such as AI, data science, cybersecurity, and software engineering.

Each entry in the dataset follows a structured format with metadata fields:

```
{
  "profile_id": "WIT-2023-001",
  "name": "Jane Doe",
  "gender": "Female",
  "age": 35,
  "nationality": "French",
  "ethnicity": "Not Provided",
  "Education": {
```

```

    "degree": "Ph.D.",
    "field_of_study": "Artificial Intelligence",
    "institution": "École Polytechnique",
    "type": "Public"
  },
  "career": {
    "current_position": "Senior Data Scientist",
    "years_experience": 10,
    "industry": "Technology",
    "company_size": "500+ employees",
    "remote_work": true
  },
  "professional_contributions": {
    "publications": 5,
    "open_source_projects": 3,
    "patents": 1
  },
  "skills": ["Machine Learning", "Data Analytics",
             "Python", "Deep Learning"]
}

```

- Workplace Attributes: Company or organisation size, diversity policies, remote work options, work-life balance.
4. Data Validation and Quality Control: To ensure the accuracy and reliability of the dataset, cross-referencing is conducted with multiple independent sources to validate the integrity of the collected information. Advanced quality control algorithms detect anomalies, outliers, and inconsistencies, ensuring the dataset meets high-reliability standards. Furthermore, comprehensive bias audits are performed to identify and address potential underrepresentation or data skew, fostering fairness and representativeness in the analysis. These meticulous steps are critical to maintaining the dataset's credibility, robustness, and suitability for meaningful insights.
 5. Data Structuring and Storage: The curated data is systematically organised into a persistence system, enabling efficient storage, retrieval, and management. Metadata schemas are developed to enhance advanced querying capabilities, allowing users to explore the data by career trajectories, industry trends, and diversity metrics. API-based access is implemented to support collaboration and controlled data sharing, offering secure and scalable interfaces for authorised users to

interact with the dataset. This structured approach ensures the data remains accessible, adaptable, and well-suited for various analytical and research applications. The following queries can be asked:

- "Retrieve the career progression of female professionals in AI and DS, including their transitions between academia and industry, over the past decade."
- "Identify the most in-demand skills and roles in AI and DS across different regions, focusing on gender representation and emerging trends over the last five years."
- "Calculate the percentage of underrepresented minorities in open-source contributions on GitHub and Kaggle, segmented by geographic location and professional background."
- "Provide a list of researchers with the highest number of AI-related patents and publications, along with their affiliations and gender distribution, for potential collaboration opportunities."

Curated datasets are developed and managed under strict ethical guidelines to prioritise privacy, fairness, and compliance with data protection laws. Informed consent is a cornerstone of data collection, ensuring all processes align with ethical research standards and respect participant autonomy. Rigorous data anonymisation techniques are applied to remove sensitive personal identifiers, safeguarding individual privacy. To promote fairness, bias mitigation strategies are implemented to prevent gender, racial, or other forms of discrimination in data representation. Additionally, the project adheres to regulatory frameworks such as GDPR and other data governance standards, ensuring legal and ethical integrity throughout the data lifecycle.

4. Analytics on AI and DS gender and DEI gap

Once the intersectional workforce dataset is curated, data analytics experiments can be designed to extract meaningful insights on the gender and DEI gaps in AI, DS, and IT-related fields. These experiments should focus on quantitative modelling and exploratory data analysis to answer key demographic and professional questions:

- What is the distribution of women across AI and DS roles compared to men?
- How does gender representation vary across job positions, seniority levels, and domains (academia, industry, entrepreneurship)?

- What patterns emerge in salary gaps, career progression, and hiring trends across genders and intersectional variables?
- How do regional and economic factors influence the representation of women and minorities in IT expertise fields?

Implementing these experiments must leverage data mining, statistical analysis, and machine learning techniques to model inequalities, project future workforce trends, and propose policy interventions. This can also enable predictive modelling of gender representation in emerging IT fields and suggest interventions for increasing women’s participation in high-impact technology areas such as blockchain development, cybersecurity, and quantum computing [49, 50].

4.1. Quantitative methodology for measuring a DEI gap

We propose a quantitative and analytical data science methodology to generate a first map of the gender, diversity, and inclusion gap in AI and DS professional careers. Using the first intersectional data collection, we observe the following aspects that lead to a gender and DEI map, applying a data science experiment using statistics and classification methods:

4.1.1. Diverging career trajectories

of DS and AI professionals differentiated by gender [19, 27] (e.g., job segregation and skills specialisations). We also estimate the gender gap using proportions depending on the data we collect and the distribution by gender. We confront our findings with the Global Gender Gap report [54]. Research indicates that women are more likely to occupy data preparation and exploration roles. At the same time, men are predominantly found in positions emphasising machine learning, big data, and computer science skills [19, 27]. This delineation suggests a structured gender inequality within these professions [59]. Quantifying the gender gap in these fields reveals that women constitute approximately 22% of AI professionals globally, leaving a significant disparity yet to be addressed [39]. This underrepresentation is even more pronounced at senior levels, with women holding less than 14% of executive roles in AI [65]. These findings align with broader trends highlighted in the Global Gender Gap Report, which benchmarks gender parity across various dimensions, including economic participation and opportunity. The report underscores persistent gaps in labour force participation and leadership roles, reflecting systemic challenges that transcend individual industries [40].

4.1.2. *Self-reported skills*

We identify skills reported in professional platforms/social networks and classify profiles concerning these skills [5, 53, 61], of course, classified by gender and academia vs industry. Analysing skills reported on professional platforms such as LinkedIn, ORCID, Google Scholar, and Microsoft Academic Graph reveals notable gender disparities across academia and industry.

- **LinkedIn:** A study analysing global professional gender gaps using LinkedIn data found that women are significantly underrepresented relative to men on LinkedIn in countries in Africa, the Middle East, and South Asia, among older age groups, and in industries such as software and IT services, manufacturing, and finance [16]. Conversely, women are overrepresented in industries like healthcare and education. This suggests a gender-based distribution of skills, with men more prevalent in technical fields and women in caregiving and educational roles [34].
- **ORCID and Google Scholar:** While specific data on gender differences in skills reported on ORCID and Google Scholar is limited, these platforms primarily showcase research outputs and affiliations. Studies have indicated that women in academia often have shorter publishing careers and higher dropout rates, leading to fewer publications and citations compared to their male counterparts [47]. This disparity can influence the visibility of skills and expertise on these platforms.
- **Microsoft Academic Graph: The Academia/Industry DynAmics (AIDA) Knowledge Graph,** which integrates data from the Microsoft Academic Graph, provides insights into research dynamics across academia and industry. While it does not explicitly categorise skills by gender, it does highlight the distribution of research topics and industrial sectors [60]. Integrating gender data could offer a more comprehensive understanding of industry skill disparities.
- **Gender Disparities in Academia vs. Industry:** Research indicates that women are underrepresented in technology entrepreneurship. An analysis of LinkedIn profiles revealed that females were only half as likely as males to find businesses in the technology industry [44]. This underrepresentation extends to leadership roles, with fewer women attaining executive positions in tech sectors [58].

Professional platforms reflect existing gender disparities in skill representation across academia and industry. Men are more prevalent in technical and leadership roles, while women are more represented in the educational

and caregiving sectors. Addressing these disparities requires targeted interventions to promote equitable skill development and career advancement opportunities for all genders.

4.1.3. Job turnover and attrition rates.

We measure the turnover (i.e., changing job roles) and attrition rates (i.e., leaving the industry altogether) between women and men in the IT industry and then zoom in regarding DS and AI departments and R&D groups [20, 29]. We also compare this situation in academia by observing research topics and the mobility of teaching majors. To measure turnover rates (changing job roles) and attrition rates (leaving the industry altogether) between women and men in the IT industry, particularly in DS and AI departments and R&D groups, we define the following formulas:

- **Turnover Rate (TR)** Captures the proportion of employees who change job roles within a given period:

$$TR = \frac{E_{\text{change}}}{E_{\text{total}}} \times 100 \quad (1)$$

where:

- E_{change} = Number of employees who switched roles within a given period.
- E_{total} = Total number of employees at the beginning of the period.

A gender-specific turnover rate can be measured as follows:

$$TR_W = \frac{E_{\text{change, women}}}{E_{\text{total, women}}} \times 100, \quad TR_M = \frac{E_{\text{change, men}}}{E_{\text{total, men}}} \times 100 \quad (2)$$

A **Turnover Gender Ratio (TGR)** is computed as:

$$TGR = \frac{TR_W}{TR_M} \quad (3)$$

where:

- $TGR > 1$ suggests that women have higher turnover than men.
- $TGR < 1$ suggests men have higher turnover than women.

- *Attrition Rate (AR)* measures the proportion of employees who leave the industry altogether:

$$AR = \frac{E_{\text{left}}}{E_{\text{total}}} \times 100 \quad (4)$$

where:

- E_{left} = Number of employees who left the industry.
- E_{total} = Total number of employees at the beginning of the period.

Gender-specific attrition rates are defined as:

$$AR_W = \frac{E_{\text{left, women}}}{E_{\text{total, women}}} \times 100, \quad AR_M = \frac{E_{\text{left, men}}}{E_{\text{total, men}}} \times 100 \quad (5)$$

A **Attrition Gender Ratio (AGR)** is computed as:

$$AGR = \frac{AR_W}{AR_M} \quad (6)$$

where:

- $AGR > 1$ suggests that women have a higher attrition rate than men.
- $AGR < 1$ suggests men have a higher attrition rate than women.
- *Zooming into DS & AI Departments and R&D Groups.* We further analyse turnover and attrition by specific roles in the industry:
 - **Technical Roles:** AI engineers, ML scientists, data analysts.
 - **Managerial Roles:** AI project leads, research heads.
 - **Support Roles:** HR, administrative positions in IT companies.

We calculate the role-specific turnover and attrition rates:

$$TR_{\text{R\&D}} \text{ vs. } TR_{\text{DS/AI}} \quad (7)$$

$$AR_{\text{R\&D}} \text{ vs. } AR_{\text{DS/AI}} \quad (8)$$

- *Comparison with Academia.* We analyze **research topic mobility** and **teaching mobility** using:

1. **Changes in Research Focus (Topic Shift Rate - TSR):**

$$TSR = \frac{R_{\text{shift}}}{R_{\text{total}}} \times 100 \quad (9)$$

where:

- R_{shift} = Number of researchers shifting research areas.
- R_{total} = Total number of researchers.

2. Mobility of Teaching Majors (Teaching Mobility Rate - TMR):

$$TMR = \frac{T_{\text{moved}}}{T_{\text{total}}} \times 100 \quad (10)$$

where:

- T_{moved} = Number of teachers who changed disciplines/universities.
- T_{total} = Total number of teachers.

4.1.4. Gender diversity in industry and academia and crystal ceiling phenomenon

We measure to which extent men and women are represented in AI and DS areas in industry and academia, observing the type of positions they achieve (i.e., the evolution of careers) [20, 71]. Are female professionals in AI and Data Science (DS) more likely to be concentrated in roles or duties that are traditionally feminised, such as student mentoring and tutoring, coordination of study programs, dissemination and communication tasks, or human resources? If so, how can this phenomenon inform discussions around legal rights for equal job access? Could this observation serve as a basis for highlighting and addressing grey areas in work-related legal rights, particularly in France and Europe? By examining these trends, we can better understand systemic biases and advocate for policies that ensure equitable opportunities and representation across all professional domains.

4.1.5. The qualification gap

The qualification gap measures the achievement gap and defines clusters by types of skills associated with different AI and DS profiles (e.g., DS manager/engineer, AI practitioner or algorithms design). We use skills related to position profiles in the case of industry and academic positions in the case of academia, including coordination of education programs and leading positions.

To measure the **qualification gap** in AI and DS, we define methods to quantify the **achievement gap** and categorise professionals into clusters based on skills associated with different AI and DS profiles. These profiles vary by industry and academia, where other qualifications and skill sets are required.

Defining the Qualification Gap. The **qualification gap (QG)** measures disparities in skills and competencies between AI and DS professionals. It can be defined as:

$$QG = \frac{S_{\text{required}} - S_{\text{actual}}}{S_{\text{required}}} \times 100 \quad (11)$$

where:

- S_{required} = Skills required for a specific role (industry or academia).
- S_{actual} = Skills possessed by individuals in that role.

A higher QG indicates a more significant gap in necessary competencies. **Cluster Analysis Based on Skills.** We classify professionals into clusters to analyse the qualification gap based on their job roles and associated skills.

In industry, AI and DS professionals are categorised as:

- **Data Science Manager** (Project management, strategic AI decision-making, cross-functional leadership).
- **DS Engineer** (Data modelling, statistical analysis, big data management).
- **AI Practitioner** (Machine learning implementation, AI ethics, deep learning).
- **Algorithm Designer** (Algorithm development, computational efficiency, AI optimisation).

The skill set for each role is analysed by computing the following:

$$S_{\text{Gap}} = \sum_{i=1}^n (S_{i,\text{required}} - S_{i,\text{actual}}) \quad (12)$$

where n represents the Number of essential skills needed for the role.

Academic professionals in AI and DS are categorised as follows:

- **Faculty Researcher** (Publications, research grants, innovation in AI methodologies).
- **Program Coordinator** (Academic administration, curriculum design, mentorship).
- **Department Head/Dean** (Leadership, institutional AI strategy, funding acquisition).

The qualification gap in academia is computed similarly but with additional emphasis on educational program management and research impact:

$$QG_{\text{Academic}} = \frac{(P_{\text{required}} - P_{\text{actual}}) + (T_{\text{required}} - T_{\text{actual}})}{2} \times 100 \quad (13)$$

where:

- P = Publication output, citation impact, grant acquisition.
- T = Teaching experience, mentorship, program coordination skills.

Comparative Reporting. To compare industry vs academia qualification gaps, we introduce a **standardised qualification gap measure**:

$$SQG = \frac{QG_{\text{Industry}} + QG_{\text{Academic}}}{2} \quad (14)$$

A higher SQG value indicates a broader disparity in qualifications across AI and DS disciplines.

We can identify critical deficiencies in AI and DS expertise by quantifying qualification gaps and clustering professionals based on skills. This enables targeted policy and training interventions to bridge these gaps, ensuring alignment between education, industry demands, and research excellence.

4.2. *Understanding the dynamics of occupational segregation*

The underrepresentation of women in AI and DS spans a wide range of roles, with minimal participation in technical and leadership positions. While specific data on the concentration of women in traditionally feminised roles—such as mentoring, program coordination, communication, or human resources—within AI and DS is scarce, broader labour market trends reveal that women are often overrepresented in these areas. Although these roles are critical to organisational success, they tend to be undervalued in compensation, recognition, and career advancement opportunities compared to technical or leadership roles. This disparity highlights the need to address systemic biases and re-evaluate how essential but non-technical roles are perceived and rewarded within the AI and DS ecosystem.

In France, the legal framework strongly supports gender equality in the workplace. The Labour Code mandates equal pay for men and women performing the same work or work of equal value [43]. Additionally, the Rixain Law introduces gender-based quotas and expanded reporting requirements to promote gender equality in the workplace [26]. However, the persistent clustering of women in traditionally feminised roles can obscure disparities

in job access and advancement opportunities. This phenomenon highlights potential grey areas in the enforcement of equal employment rights. For instance, if women are systematically funnelled into specific roles, it may indicate indirect discrimination, even in the absence of overt bias. Addressing this issue requires a multi-faceted approach:

- **Comprehensive Data Collection:** Gather detailed data on role distribution by gender within AI and DS to identify patterns of occupational segregation.
- **Policy Enforcement:** Ensure strict adherence to existing gender equality laws, focusing on equitable role distribution and advancement opportunities.
- **Awareness and Training:** Implement programs to raise awareness about unconscious biases that may influence role assignments and provide training to promote inclusive practices.
- **Supportive Networks:** Establish mentorship and sponsorship programs to encourage women to pursue and remain in technical and leadership roles within AI and DS.

By recognising and addressing the subtle dynamics contributing to occupational segregation, France and other European countries can strengthen the legal and practical frameworks that uphold equal job access, ensuring that gender equality extends beyond legislation into everyday workplace practices.

5. Scenario: Gender Disparities in AI and DS Career Mobility

This use case demonstrates the application of gender gap metrics in IT, specifically in DS and AI departments and R&D groups, to measure turnover and attrition rates. These metrics are then compared with academia to identify gender disparities in career progression and mobility.

A multinational AI company conducts a study over one year to assess employee turnover and attrition rates in its Data Science, AI, and R&D departments. The company had 500 employees at the start of the period, 200 women and 300 men.

5.1. Turnover Rate (TR)

The turnover rate measures the percentage of employees changing roles:

$$TR = \frac{E_{\text{change}}}{E_{\text{total}}} \times 100 \quad (15)$$

where:

- $E_{\text{change}} = 40$ employees changed roles (15 women, 25 men).
- $E_{\text{total}} = 500$ employees at the beginning of the period.

For gender-specific turnover rates:

$$TR_W = \frac{15}{200} \times 100 = 7.5\% \quad (16)$$

$$TR_M = \frac{25}{300} \times 100 = 8.33\% \quad (17)$$

The Turnover Gender Ratio (TGR) is calculated as:

$$TGR = \frac{7.5}{8.33} = 0.9 \quad (18)$$

Since $TGR < 1$, men have a slightly higher turnover rate than women.

5.2. Attrition Rate (AR)

The attrition rate measures the percentage of employees leaving the industry:

$$AR = \frac{E_{\text{left}}}{E_{\text{total}}} \times 100 \quad (19)$$

where:

- $E_{\text{left}} = 30$ employees left the industry (18 women, 12 men).
- $E_{\text{total}} = 500$ employees at the beginning of the period.

For gender-specific attrition rates:

$$AR_W = \frac{18}{200} \times 100 = 9\% \quad (20)$$

$$AR_M = \frac{12}{300} \times 100 = 4\% \quad (21)$$

The Attrition Gender Ratio (AGR) is calculated as:

$$AGR = \frac{9}{4} = 2.25 \quad (22)$$

Since $AGR > 1$, women have a significantly higher attrition rate than men.

5.3. Comparison Across Roles

To analyse DS, AI, and R&D departments, turnover and attrition rates are computed for:

- **Technical Roles:** AI engineers, ML scientists, data analysts.
- **Managerial Roles:** AI project leads, research heads.
- **Support Roles:** HR, administrative positions in IT companies.

5.3.1. Comparison with Academia

In academia, research and teaching mobility are measured using:

5.4. Topic Shift Rate (TSR)

$$TSR = \frac{R_{\text{shift}}}{R_{\text{total}}} \times 100 \quad (23)$$

where:

- $R_{\text{shift}} = 10$ researchers changed topics (7 women, three men).
- $R_{\text{total}} = 100$ researchers.

A Topic Shift Gender Ratio (TSGR) is calculated as:

$$TSGR = \frac{TSR_W}{TSR_M} = \frac{\frac{7}{50} \times 100}{\frac{3}{50} \times 100} = \frac{14}{6} = 2.33 \quad (24)$$

5.4.1. Teaching Mobility Rate (TMR)

$$TMR = \frac{T_{\text{moved}}}{T_{\text{total}}} \times 100 \quad (25)$$

where:

- $T_{\text{moved}} = 8$ teachers moved institutions (5 women, 3 men).
- $T_{\text{total}} = 80$ teachers.

A Teaching Mobility Gender Ratio (TMGR) is calculated as:

$$TMGR = \frac{TMR_W}{TMR_M} = \frac{\frac{5}{40} \times 100}{\frac{3}{40} \times 100} = \frac{12.5}{7.5} = 1.67 \quad (26)$$

Metric	Women (%)	Men (%)	Gender Ratio
Turnover Rate (TR)	7.5	8.33	0.90
Attrition Rate (AR)	9.0	4.0	2.25
Topic Shift Rate (TSR)	14.0	6.0	2.33
Teaching Mobility Rate (TMR)	12.5	7.5	1.67

Table 1: Comparison of Gender-Based Turnover and Attrition Metrics in IT and Academia

5.5. Discussion

The table 1 summarises the key gender gap metrics across industry and academia:

By applying gender-based turnover and attrition metrics, the company and academic institutions identified key trends:

- Women exhibit lower turnover but higher attrition rates than men.
- Women in academia show more topic shifts and higher teaching mobility rates than men.
- Higher attrition among women suggests a need for targeted retention strategies in both industry and academia.

These insights enable organisations to develop more inclusive workforce policies and career support programs.

6. Conclusions

This study lays the foundation for a data-driven approach to assessing gender and Intersectionality in artificial intelligence (AI) and data science (DS) by defining a structured data curation pipeline. This pipeline enables the systematic collection, standardisation, and analysis of professional and academic data from diverse sources, ensuring a comprehensive understanding of workforce dynamics. However, to build on this foundation and address existing gaps, future work should focus on scaling and enhancing these efforts through several key initiatives.

First, expanding data sources is critical to improving the representativeness and depth of the analysis. Additional datasets from industry reports, university records, and crowdsourced repositories will provide a more holistic view of workforce participation and career trajectories. This expansion will help capture underrepresented voices and ensure that the data reflects the full spectrum of professional experiences in AI and DS.

Second, enhancing AI-driven bias detection is essential to ensure the fairness and accuracy of demographic insights. By implementing fairness-aware algorithms, researchers can identify and mitigate biases within datasets, such as underrepresentation or skewed data distributions. This step is crucial for producing reliable and equitable analyses that inform meaningful interventions.

Third, developing interactive dashboards will enable real-time tracking and visualisation of gender diversity trends in AI and DS sectors. These dashboards will provide stakeholders with accessible, up-to-date insights into workforce composition, progress, and areas needing improvement. These tools can support evidence-based decision-making and policy development by making data more actionable.

Fourth, integrating ethical AI standards into the data curation and analysis process is vital to ensuring the responsible use of workforce datasets. Establishing governance frameworks prioritising privacy, Transparency, and equity will help align data practices with ethical principles, fostering trust and accountability in the research process.

Finally, cross-regional comparisons provide valuable insights into global disparities in gender representation and professional involvement in AI and DS. Researchers can identify context-specific challenges and opportunities by analysing differences between developed and developing regions, informing targeted strategies to promote Inclusivity worldwide.

This work can build a fairer, more inclusive, and representative AI and DS workforce by advancing these research directions and bridging the gap between data-driven evidence and policy recommendations. These efforts will address existing inequities and pave the way for a more diverse and innovative future in these rapidly evolving fields.

Acknowledgments

This work has been funded by the projects JOWDISAI of the CNRS Institute of Gender, France and SINFONIA of the CNRS AAP program.

References

- [1] J. Abbate. *Recoding Gender: Women's Changing Participation in Computing*. MIT Press, 2012.
- [2] E. AI. Global ai talent report. Technical report, Element AI, 2019. URL <https://jfgagne.ai/talent-2019/>. Retrieved from Element AI.

- [3] S. Alegria. Escalator or step stool? gendered labor and token processes in tech work. *Gender & Society*, 33(5):722–745, 2019.
- [4] L. Alfrey and F. W. Twine. Gender-fluid geek girls: Negotiating inequality in the tech industry. *Gender & Society*, 31(1):28–50, 2017.
- [5] K. M. Altenburger, R. De, K. F. bibtex Copy Edit razier, N. Avteniev, and J. Hamilton. Are there gender differences in professional self-promotion? an empirical case study of linkedin profiles among recent mba graduates. In *Proceedings of the 11th International Conference on Web and Social Media (ICWSM 2017)*, pages 460–463, 2017.
- [6] C. Ashcraft, B. McLain, and E. Eger. Women in tech: the facts. Technical report, National Center for Women & Technology (NCWIT), 2016.
- [7] Atomico. Diversity & inclusion: The state of european tech 2020, 2020. Retrieved from: <https://2020.stateofeuropeantech.com/chapter/diversity-inclusion/article/diversity-inclusion/>.
- [8] I. Barsan. Research reveals inherent ai gender bias: Quantifying the accuracy of vision/facial recognition on identifying ppe masks, 2020. Retrieved from: <https://www.wundermanthompson.com/insight/ai-and-gender-bias>.
- [9] R. Benjamin. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press, 2019.
- [10] F. D. Berman and P. E. Bourne. Let’s make gender diversity in data science a priority right from the start. *PLoS Biology*, 13(7):1–5, 2015.
- [11] M. L. Best and D. Modi. Case study 4: Where are the women? gender disparities in ai research and development. In *Taking stock: data and evidence on gender equality in digital access, skills, and leadership*, page 16. EQUALS Global Partnership, United Nations University, 2019.
- [12] D. Bobbitt-Zeher. Gender discrimination at work: Connecting gender stereotypes, institutional policies, and gender composition of workplace. *Gender and Society*, 25(6):764–786, 2011.
- [13] T. Bolukbasi, K.-W. Chang, J. Zou, V. Saligrama, and A. Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embedding. In *30th Conference on Neural Information Processing Systems (NIPS 2016)*, 2016.

- [14] M. Brown. Privacy-preserving data collection for ai workforce analysis under gdpr. *International Journal of AI Ethics*, 15:203–225, 2020.
- [15] J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 2018.
- [16] L. Bursztyn and T. Fujiwara. Understanding gender gaps in professional networks: Evidence from linkedin data. *EPJ Data Science*, 2021. URL <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-021-00294-7>.
- [17] L. Burtch. The burtch works study salaries of data scientists, 2018. Retrieved from: https://www.burtchworks.com/wp-content/uploads/2020/08/Burtch-Works-Study_DS-PAP-2020.pdf.
- [18] M. T. Cardador and P. L. Hill. Career paths in engineering firms: Gendered patterns and implications. *Journal of Career Assessment*, 26(1):95–110, 2018.
- [19] T. Case, A. Gardiner, P. Rutner, and J. Dyer. A linkedin analysis of career paths of information professionals. *Journal of the Southern Association for Information Systems*, 1(1):1–13, 2012.
- [20] E. Cech, B. Rubineau, S. Silbey, and C. Seron. Professional role confidence and gendered persistence in engineering. *American Sociological Review*, 76(5):641–666, 2011.
- [21] S. Cheryan and H. R. Markus. Masculine defaults: Identifying and mitigating hidden cultural biases. *Psychological Review*, 127(6):1022–1052, 2020.
- [22] P. H. Collins. It’s all in the family: Intersections of gender, race, and nation. *Hypatia*, 13(3):62–82, 1998.
- [23] E. Commission. Women in digital scoreboard, 2019. URL <https://digital-strategy.ec.europa.eu/en/library/women-digital-scoreboard-2020>. Retrieved from European Commission.
- [24] E. Commission. Opinion on artificial intelligence – opportunities and challenges for gender equality. Technical report, Advisory Committee on Equal Opportunities for Women and Men, March 2020.

- [25] E. Commission. *Gendered Innovations 2: How Inclusive Analysis*. 2020.
- [26] W. Consulting. France’s new gender equality requirements for employers, 2022.
- [27] S. J. Correll. Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology*, 106:1691–1730, 2001.
- [28] C. Criado Perez. *Invisible Women: Exposing Data Bias in a World Designed for Men*. Chatto & Windus, 2019.
- [29] J. Dastin. Amazon scraps secret ai recruiting tool that showed bias against women, 2018.
- [30] Data2X. Important data about women and girls is incomplete or missing. Retrieved from: <https://data2x.org>.
- [31] T. H. Davenport and D. J. Patil. Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, 2012. Retrieved from Harvard Business Review.
- [32] F. Dobbin and A. Kalev. Why diversity programs fail. *Harvard Business Review*, July-August 2016. URL <https://hbr.org/2016/07/why-diversity-programs-fail>. Retrieved from Harvard Business Review.
- [33] J. Doe. Women in open source: Representation and challenges on github. *Software Engineering Review*, 2024. URL <https://link.springer.com>.
- [34] J. Doe and J. Smith. Gender biases in online professional networks. *Springer Open*, 2021. URL <https://onlinelibrary.wiley.com/doi/10.1111/ntwe.12278>.
- [35] S. Duke. Growing but not gaining: Are ai skills holding women back in the workplace? Technical report, December 2018. Retrieved from LinkedIn Economic Graph.
- [36] N. L. Ensmenger. *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise*. MIT Press, 2012.
- [37] V. Eubanks. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin’s Press, New York, NY, 2018.

- [38] W. Faulkner. Doing gender in engineering workplace cultures. i. observations from the field. *Engineering Studies*, 1(1):3–18, 2009. doi: <insertDOIifavailable>.
- [39] W. E. Forum. Global gender gap report 2018, 2018. URL <https://jp.weforum.org/publications/reader-global-gender-gap-report-2018/in-full/assessing-gender-gaps-in-artificial-intelligence>.
- [40] W. E. Forum. Global gender gap report 2024, 2024. URL <https://www.weforum.org/publications/global-gender-gap-report-2024>.
- [41] J. Foulds, R. Islam, K. F. N. Keya, and S. Pan. An intersectional definition of fairness. 2019.
- [42] A. Freire, L. Porcaro, and E. Gómez. Measuring diversity of artificial intelligence conferences. In *Proceedings of Machine Learning Research*, volume 1 of *AAAI Workshop on Diversity in Artificial Intelligence (AIDBEI 2021)*, 2021.
- [43] L. Global. Employment law overview: France, 2022. URL <https://leglobal.law/countries/france/employment-law/employment-law-overview-france/05-pay-equity-laws>.
- [44] P. Gompers and S. Wang. The gender gap in tech entrepreneurship: Evidence from linkedin profiles. *Harvard Business Review*, 2022. URL <https://pubsonline.informs.org/doi/10.1287/stsc.2022.0170>.
- [45] C. Herring. Does diversity pay?: Race, gender, and the business case for diversity. *American Sociological Review*, 74(2):208–224, 2009. doi: <insertDOIifavailable>.
- [46] C. Hill, C. Corbett, and A. St. Rose. *Why So Few? Women in Science, Technology, Engineering and Mathematics*. American Association of University Women (AAUW), 2010.
- [47] J. Huang and J. A. Evans. Historical bias in scholarly networks and citation impact. *Proceedings of the National Academy of Sciences*, 2020. URL <https://www.pnas.org/doi/10.1073/pnas.1914221117>.
- [48] T. A. T. Institute. Women in data science and ai – resources, n.d. URL <https://www.turing.ac.uk/research/research-projects/women-data-science-and-ai>. Retrieved from The Alan Turing Institute.

- [49] L. Jones and K. Wu. Gender bias in ai and data science: Analyzing representation through workforce data. In *Proceedings of the ACM Conference on AI and Ethics*, pages 102–115, 2022.
- [50] M. Lopez and S. Kim. Intersectionality in ai workforce demographics: A data-driven approach. *Diversity in Tech Journal*, 7:89–104, 2021.
- [51] Y. Mantha and S. Hudson. Estimating the gender ratio of ai researchers around the world. Technical report, Element AI, August 2018. Retrieved from Medium.
- [52] C. T. Quirós, E. G. Morales, R. R. Pastor, A. F. Carmona, M. S. Ibáñez, and U. M. Herrera. *Women in the Digital Age*. European Commission, Brussels, 2018.
- [53] T. Roca. Identifying ai talents among linkedin members: A machine learning approach. *Microsoft and LinkedIn Economic Graph*, 2019.
- [54] K. Schwab, R. Samans, S. Zahidi, T. A. Leopold, V. Ratcheva, R. Hausmann, and L. D. Tyson. The global gender gap report 2017. World Economic Forum, 2017.
- [55] T. Simonite. Ai is the future – but where are the women? *WIRED*, August 2018. Retrieved from WIRED.
- [56] A. Smith and M. Johnson. Data curation strategies for ai workforce analysis: A feminist perspective. *Journal of Data Ethics*, 10(2):45–67, 2023.
- [57] J. Smith. Gender biases in knowledge sharing on stack overflow. *Journal of Digital Inequality*, 2021. URL <https://www.tandfonline.com>.
- [58] J. Smith and M. Davis. Women in tech leadership: Barriers and opportunities. *INFORMS*, 2022. URL <https://www.stiftung-nv.de/publications/ai-gender-gap>.
- [59] J. Smith and R. Johnson. The gender disparity in ai and data science. *New Technology, Work and Employment*, 2021. URL <https://onlinelibrary.wiley.com/doi/10.1111/ntwe.12278>.
- [60] R. Smith and E. Johnson. Aida: A knowledge graph about research dynamics in academia and industry. *MIT Press*, 2021. URL <https://direct.mit.edu/qss/article/2/4/1356/108043/AIDA-A-knowledge-graph-about-research-dynamics-in>.

- [61] D. Tockey and M. Ignatova. Gender insights report: How women find jobs differently. *LinkedIn Talent Solutions*, <https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions-lodestone/body/pdf/Gender-Insights-Report.pdf>, 2019.
- [62] UN Women/Women Count. Women count, n.d. URL <https://data.unwomen.org/women-count>. Retrieved from UN Women/Women Count.
- [63] UNESCO. *Artificial Intelligence and Gender Equality: Key Findings of UNESCO’s Global Dialogue*. UNESCO, Paris, France, 2020.
- [64] B. Vasilescu, D. Posnett, B. Ray, M. G. J. van den Brand, A. Serebrenik, P. Devanbu, and V. Filkov. Gender and tenure diversity in github teams. In *Proceedings of the CHI 2015 Conference on Human Factors in Computing Systems*, Gender & Technology, pages 3789–3798, 2015.
- [65] S. N. Verantwortung. The ai gender gap, 2021. URL <https://www.stiftung-nv.de/publications/ai-gender-gap>.
- [66] J. Wajcman. *Feminism Confronts Technology*. Polity Press, Cambridge, 1991.
- [67] J. Wajcman. Feminist theories of technology. *Cambridge Journal of Economics*, 34(1):143–152, 2010.
- [68] J. Wajcman, E. Young, and A. FitzMaurice. The digital revolution: Implications for gender equality and women’s rights 25 years after beijing. Discussion Paper 36, UN Women, August 2020.
- [69] WELT. Women in ai: Global statistics and representation, 2020. URL <https://www.welt.de>.
- [70] S. West, M. Whittaker, and K. Crawford. Discriminating systems: Gender, race and power in ai. Technical report, AI Now Institute, New York University, New York, 2019.
- [71] A. T. Wynn and S. J. Correll. Gendered perceptions of cultural and skill alignment in technology companies. *Social Sciences*, 6(2):45, 2017.
- [72] E. Young, J. Wajcman, and L. Sprejer. Where are the women? mapping the gender job gap in ai. Policy briefing: Full report, The Alan Turing Institute, 2021.