Imperial College London

Department of Life Sciences

Silwood Park Campus

Quantifying the gender gap in UK STEM funding

Author: Xuan Wang
CID: 02150780
email: xuan.wang22@imperial.ac.uk

Supervisor:
Samraat Pawar
s.pawar@imperial.ac.uk

A thesis submitted in partial fulfilment for the degree of
Master of Science at Imperial College London
Submitted for MSc Computational Methods in Ecology and Evolution
August 2024

Declaration

For this study, I utilised two primary datasets: one detailing staff numbers sourced from the Higher Education Statistics Agency (HESA) official website, and another containing information on UK funding projects within the Science, Technology, Engineering, and Mathematics (STEM) domain, obtained from the UK Research and Innovation (UKRI) website. I used the UKRI dataset shared by my colleague, Flavia Trigo, which can be assessed on Onedrive; For the analysis, I downloaded the HESA data from the official website. The (open source) code for the topic modelling was provided by Flavia Trigo, which is public and can be retrieved from Github. I independently developed the code for the entire analysis, which involved processing the outcomes from the topic analysis, creating visualizations, and conducting hypothesis tests. Under the invaluable guidance of my supervisor, Samraat Pawar, I received precious advice on statistical analyses, direction on the writing process and insights on the types of results to showcase.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Samraat Pawar, Imperial College London for invaluable guidance, expertise and support throughout the duration of the project.

Contents

List of Figures

Abstract

Advancements in the fields of science, technology, engineering, and mathematics (STEM) have profoundly transformed the global landscape. In response, the UK government has introduced numerous strategies to increase funding in STEM, with the aim of fostering innovation and research. However, the equitable distribution of these benefits for different groups of people remains under scrutiny, which could directly affect the incentive of researchers and impact scientific innovation. My research investigates gender disparities in the amount of funding and proportion of each gender getting funded within STEM disciplines, using a machine learning approach, Natural Language Processing (NLP), which has not been previously applied to address this issue, thus exploring this critical gap. I used a raw data set comprising 107,760 projects for this investigation, where the projects were classified by topic analysis and ChatGPT. The findings were then compared and analysed among various STEM categories in terms of the average amount of funding and the funding ratio of each gender. Although the results confirmed the presence of gender disparity in both the average amount of STEM research funding and the funding ratio for each gender, they also indicated a general decline in bias over time. Among the STEM categories, chemistry stands as an exception, in which the gender gap has increased in recent years. The overall pattern of gender disparity may stem from the historically accumulated advantages of males, including biases in recruitment and promotion processes, differences in life goals, and the absence of specific equality policies, particularly in research funding. To tackle this issue, a comprehensive approach is needed to strengthen the academic structure, including efforts to promote diversity in university employment and revising the criteria for funding applications.

1 Introduction

Objectivity is fundamental in academic practice to optimally assess one's ability and research findings [Handley et al. (2015)]. Nevertheless, there is still evidence of gender disparity in many aspects worldwide, where the difference is more prominent among academics working in Science, Technology, Engineering and Mathematics (STEM) [García-González et al. (2019)]. It has been established that women represent 53% of bachelor's and master's degree graduates globally, but this proportion decreases to 43% at the PhD level. Despite this, women hold only 28% of research positions [García-González et al. (2019)]. This difference is more noticeable in senior-level positions. The She Figures 2021 data, which studies the gender situation in European countries, has shown that only 26.2% of the grade-A highest academic positions were held by women [Commission (2020)]; female researchers in Ireland accounted for 33% of full professors according to the data published by Dublin City University (DCU) in 2019 [Hosseini and Sharifzad (2021)]; and this number drops to 25% full professors in United Kingdom (UK) by the data published by Higher Education Statistics Agency (HESA) in 2018, and 32% of university governing board members [Thelwall et al. (2020)].

A less-prominent position for females in the academic field could lead to an unbalanced study environment, exacerbating the shortage of skilled workers in STEM and also adversely affecting the situation of female workers [Verdugo-Castro et al. (2022)]. The UK government Science and Technology Committee reported in 2014, on women's scientific careers in the UK, that

> *"...the problems and solutions have long been identified, yet not enough is being done to improve the situation"*

White and Smith (2022). This unbalanced environment could result in fewer measures to tackle gender-related problems García-González et al. (2019), as previous literature has shown that studies by male researchers are generally less likely to judge gender discrimination than those by females Handley et al. (2015). Results within such a biased academic environment can be irrelevant or even harmful to women Cislak et al. (2018). [White and Smith (2022)]. This biased environment could result in fewer measures to tackle gender-related problems [García-González et al. (2019)], as the previous literature has shown that studies by male researchers are generally less likely to judge gender discrimination than those by females [Handley et al. (2015)]. Results within such a biased academic environment can be irrelevant, or even harmful, to women [Cislak et al. (2018)].

STEM studies, which are fundamental for the resolution of pressing global issues such as climate change and human health, are known to have a considerable gender gap around the world [Bellotto Trigo et al. (2023); García-Holgado et al. (2019a)]. Although there is a growing body of research addressing the diversity gap across multiple facets of the STEM domain (employment, salary, etc.) to ensure the payoffs from STEM are equitable, the impartiality of these studies remains subject to scrutiny due to the current gender difference in researchers. Recent labour market data in 2022/23 indicates that women comprised only about 26% of the STEM workforce in the UK, suggesting that equal representation may not be achieved until 2070 at the current rate of growth [Women (2023)]. This imbalance could exacerbate a cumulative bias known as the "Mathew Effect"[Jebsen et al. (2020)]. For instance, since women are less likely than men to secure positions in higher-ranking institutions, they may face additional challenges when applying for funding [Cruz-Castro et al. (2022)]. Consequently, more research is necessary to address the gender gap in STEM in a comprehensive way.

Currently, there has been an increasing number of studies that have focused on the gender gap in STEM, while the gap still exists from the aspect of STEM funding, and the methodology differs. The review by Verdugo-Castro et al. (2022) found that the number of related literature reached a maximum in 2017 and 2018 [Verdugo-Castro et al. (2022)]. The researchers tried to find the possible reason for the gender differences. The study by Kang et al. (2019) in Finland found the existence of clear gender differences among children's interests regarding science subjects, while teachers are an important agent that may influence their future path. Another common point is the challenges related to pregnancy and caring responsibilities, which could significantly disadvantage female researchers when returning after a career break [Craig and Powell (2011)]. However, there are differences among the methods used in these studies. Among existing studies, quantitative methods are still the most popular analysis method [Verdugo-Castro et al. (2022)]. The study by Delaney and Devereux (2019) applied regression analysis to find differences in the choices and acceptance of STEM courses by gender in Ireland; Kube et al. (2024) applied group concept mapping in recent research to see the impact of gender bias in STEM education in Germany, and then aggregated the data through multidimensional scaling and hierarchical cluster analysis; Spatial visualisation is also a popular methodology among present meta-data research in this field [Lee et al. (2019)]. However, to our knowledge, machine learning methods have not yet been applied in this field. Furthermore, a majority of current studies focus on the aspect of STEM education and employment to find the gender gap, while comparatively less attention is paid to the funding to STEM, which is indeed an essential factor influencing the incentive of female researchers to produce research.

As one of the top global corporate Research & Development (R&D) investors[WIPO (2022)], the UK government announced a £39.8 billion R&D budget for 2022-2025 to stimulate the power in science development [UK Space Agency (2024)]. However, data from the Engineering and Physical Sciences Research Council (EPSRC) in 2016-2017 showed that £944 million of funding was awarded to male applicants while only £69 million went to female applicants [Jebsen et al. (2020)]. An unequal opportunity for research funding would reduce the incentive and access to institutional resources [Cruz-Castro et al. (2022)]. Therefore, in this study, my aim is to determine whether there is a gender bias in STEM from the aspect of research funding, specifically in the UK. For this purpose, I analyze the general distribution of gender ratios and subsequently compare them across different categories and institutions. My primary focus will be to answer the following questions:

- From 2015 to 2022, how has the funding trend for each gender evolved? Is there a noticeable disparity?
- If there is a gender disparity, what could be the underlying cause?
- Is there any implication of the bias?

To address the first question, this report will assess the overall pattern of funding amount and ratio over time for each gender, and hypothesis testing will be employed for evidence. The subsequent questions will be explored by comparing the trend among different STEM categories, considering recent policies and previous literature.

2 Methods

In my study, I used a machine learning method called topic analysis on the entire data set to identify the subject of each project. One of the main advantages of this technique is the efficiency and accuracy of topic analysis when examining large and complex datasets. At the same time, topic analysis usually provides a more objective result, as it can identify underlying themes from texts independently of potentially biased perspectives, which may exist in keywords provided by applicants [Hagen (2018)]. Subsequently, the results after the application of topic modelling were classified using ChatGPT due to its consistency, efficiency, and reduction of human error compared to manual classification. This classified data will be used for further analysis. After identifying the gender of the lead author for each project, the final step involves quantifying the overall gender proportions and comparing these across different STEM categories over time.

To account for all biases, including before the application stage, the ratio is calculated by dividing the funded males or females by the total number of staff in each classification for comparison.

2.1 Data acquisition and characteristics

I used three datasets. The first consisted of title abstracts from 107,760 projects collected from UKRI, provided by our colleague [Bellotto Trigo et al. (2023)]. The second dataset contained relevant project information, which I obtained from UKRI, including the leading institution, the applicant's name, the start and end date of the project, etc. These two raw datasets were merged for further analysis.

The third data set is on the information of the entire academic staff in the UK STEM area, sourced from the website of the Higher Education Statistics Agency (HESA). This includes the gender data of staff in various academic fields. As the HESA data were available only from 2015 to 2022 at the time of our study was taken, we limited our analysis to title abstract data within this period.

2.2 Procedures

2.2.1 Data processing

The first step of our study is the pre-processing of the metadata. The raw dataset of the title abstracts is cleaned by several approaches, including filtering of STEM funding bodies and the selection of valuable columns. The information in the clean UK data includes the project ID, funding body, the lead institution, the applicant's name, the date of the project, and the funding amount. In addition, the title abstracts are tokenized and filtered by dropping the rows where the number of tokens is less than 9, selecting only the data that provide meaningful information.

After completing the pre-processing of the raw metadata, I used Natural Language Processing (NLP), particularly topic analysis, on the cleaned data for the identification of topics. Topic analysis is a machine learning technique to identify topics by finding common themes in vast amounts of text [Bellotto Trigo et al. (2023)]. Specifically, the Latent Dirichlet Allocation (LDA) model was applied to our cleaned data to find the underlying topic of each project in our dataset. As an example of topic analysis, LDA can detect the underlying topics in a collection of documents and then determine how likely they belong to each topic by generating a likelihood distribution result [Wikipedia (2023)]. We applied LDA models with 50 to 200 topics in 25 increments to the data, and calculated the perplexity value for each of the models with different numbers of topics to determine the best-fitted model. The model with 200 topics is finally selected due to its lowest perplexity value compared to the others, indicating better predictive performance.

2.2.2 Analysis

Upon applying LDA to the clean data, I obtained the conditional probability distribution of each project belonging to each of the 200 topics. I choose the topic with the highest probability as the corresponding topic for each project. The next step is the classification of the issues. The results generated from LDA involve a comprehensive collection of keywords for each topic. To enhance efficiency and accuracy, I use the advanced machine learning tool, ChatGPT, to sort the 200 topics into the following STEM categories according to their keywords: Veterinary Science; Agriculture, Forestry & Food Science; General Engineering; Chemical Engineering; Mineral, Metallurgy & Materials Engineering; Civil Engineering; Electrical, Electronic & Computer Engineering; Mechanical, Aero & Production Engineering; IT, Systems Sciences & Computer Software Engineering; Earth, Marine & Environmental Sciences; Biosciences; Physics; Chemistry; Mathematics.

To determine the gender for each project, I applied a Python package, Gender Guessor, to the leading applicant of each project in the clean data to distinguish their gender based on their names. The output of this Python package includes Male, Female, Mostly male, Mostly female, Unknown, and Androgynous. In this study, both "Male" and "Mostly male" outputs are classified as male, while "Female" and "Mostly female" are treated as female, and the other results are not considered due to the higher uncertainty and inaccuracy. The gender information is combined with the classification of each project. The combined dataset is then used for further analysis, aiming to address the following segments:

1. Comparison of the average funding amount in 2015-2022

To detect the existence of a gender gap, the first step is to examine the distribution of funding amounts between genders. Although the government has announced significant funding for STEM, the proportion allocated to females remains unknown. It is crucial to determine whether each gender has equal access to funding. Therefore, the initial task of this study is to assess the average amount of funding for each gender. This involves identifying the total amount of funding during the period and dividing it by the number of males and females funded, respectively. I will then compare the average funding distributed to each gender in the UK from 2015 to 2022 in various categories.

2. Tracking gender-based funding trends in each STEM category

Despite recent policy initiatives aimed at promoting STEM research through increased funding, it remains unclear whether these measures will effectively increase innovation among female researchers. Therefore, after identifying the general pattern, another objective of this study is to track the funding trends for each gender within each classification during the period and to determine whether there is a gender disparity and whether current policies have positively impacted it. The average funding identified in the first step will be compared over time.

3. Comparison of funding bias ratio across different categories.

The next phase involves comparing the degree of bias to identify which area may contribute the most to the potential disparity. A funding ratio for male and female researchers is computed by dividing the number of funded researchers within each gender by the total number of staff members of the same gender. In the above steps, we can detect the potential bias of the funded researchers. However, to understand the underlying reasons for the bias, it is essential to consider the bias at all stages, including before and after the application process - the potential reason for each gender satisfying the application requirement, the chance of each gender applied for funding that finally getting funded, as well as the general pattern for funded researhcers that would be addressed in the above questions, should all be considered. Therefore, this explains why my study uses the total number of academic staff to calculate the gender ratio. These ratios for each gender are compared over the years. Finally, the degree of bias across all categories is compared by calculating the difference in the gender ratios for each gender, referred to as the bias ratio. A higher value indicates a greater degree of gender bias.

3 Results

3.1 Average funding trends over 2015-2022

Figure 1 shows that there is a large gender gap in most engineering-related subjects, except Chemical Engineering, where the gap is rather small. In general, the average funding amount for male researchers is greater than that for female researchers, except in the fields of Physics and Veterinary Science, where the average funding amount for females is larger than that for males. The findings in Figure 2 closely align with the observations from Figure 1. In particular, within the field of Physics, female researchers obtained a higher average funding compared to male researchers, primarily because of the markedly increased funding for women in 2020. During other years, males typically received more funding, or the funding amounts were nearly the same. In order to test for the gender gap, hypothesis testing is required. The figure below shows the distribution of the funding amount for female researchers and male researchers. We can observe that both distributions are left skewed. Given that the observations are independent, we applied Mann-Whitney U test in this study. The test evaluated the whether the mean funding for male and female researchers is equal.

The result of the Mann-Whitney U test shows a p-value of 1.181e-05, which is smaller than a 5% significant level. This result indicates a highly statistically significant difference between the mean

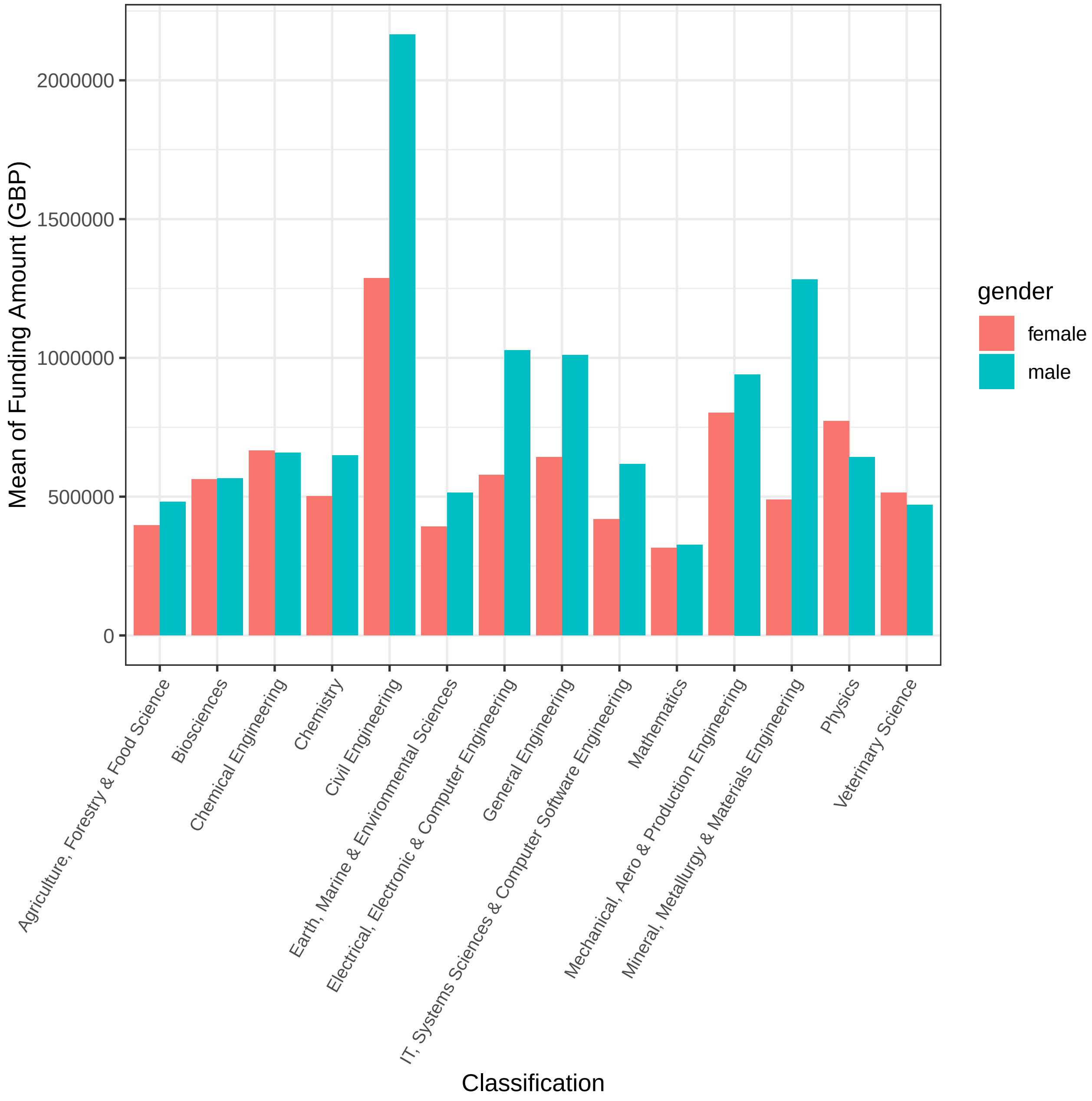

Figure 1: Average research funding allocated to each gender during the period measured. The plot shows a gap in most of the STEM subjects, and that the funding amount for male researchers is generally higher.

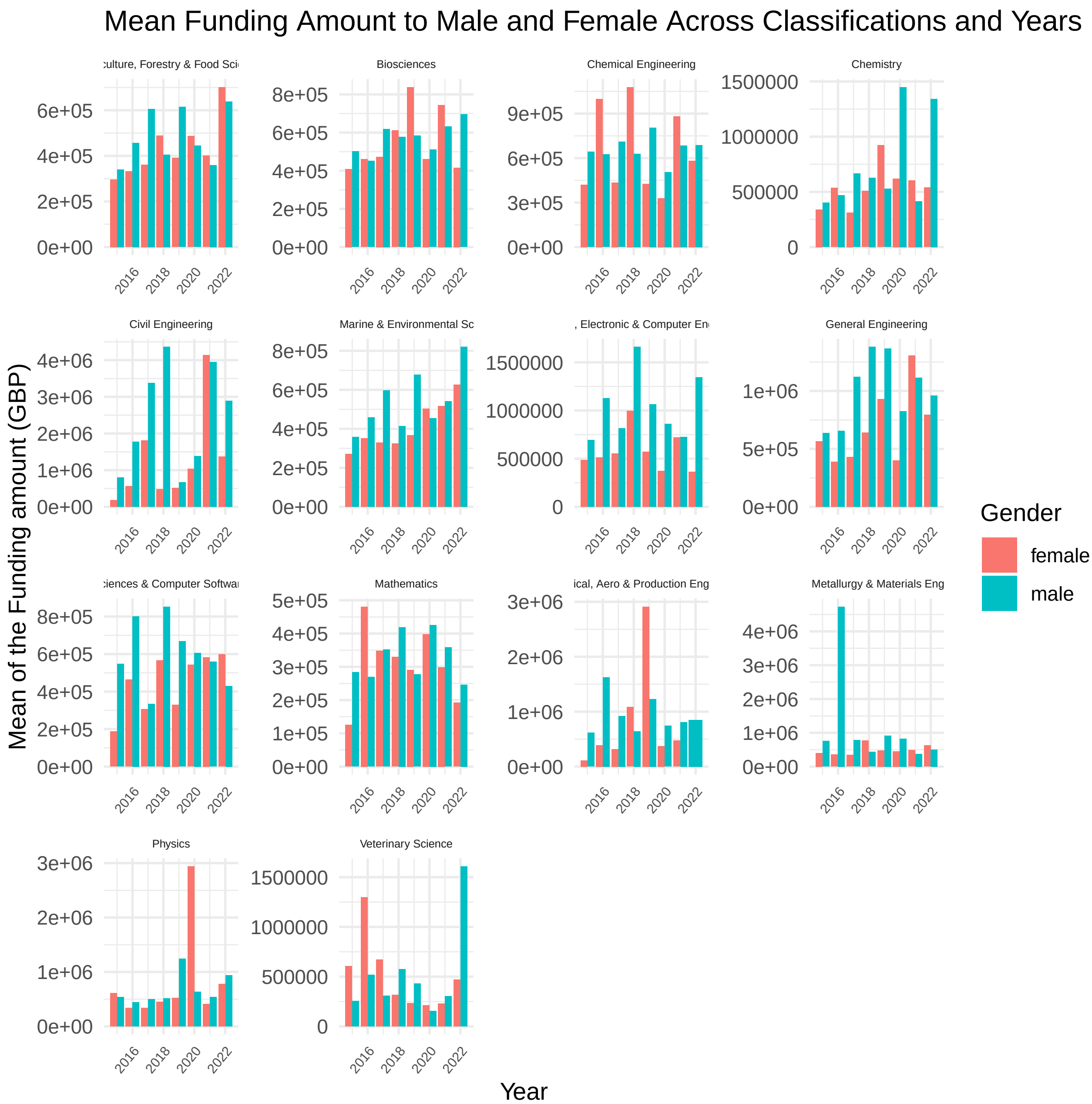

Figure 2: Mean funding allocated to each gender over time. We can observe that the gender gap in Chemistry increased over time; for Veterinary Science, the funding amount to female researchers was initially higher than male, and became lower in recent years. In most engineering-related subjects, there is a noticeable gender gap over time.

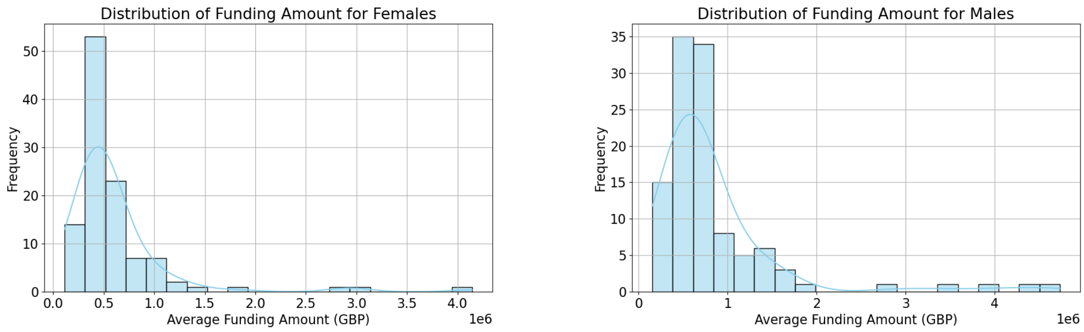

Figure 3: The distribution of average funding amount. The plots are left-skewed, meaning that a parametric test may not be a good choice.

funding for male and female researchers, providing strong evident to reject the null hypothesis and to the existence of gender gap between the mean funding amount.

3.2 Comparison of the proportion of each gender getting funded

3.2.1 Comparison among classifications

As displayed in Figures 4 and 5, the results show a significant gender gap in the research funding of Biosciences, where the gap almost persists over time; although the extent of gender bias overall is the greatest in Agriculture, Forestry & Food Science, the disparity has been considered to have decreased in the last few years. In contrast, there are several categories where the proportion of both genders is nearly the same over time, including General Engineering, Civil Engineering, and IT, Systems Sciences & Computer Software Engineering. In most classifications, there is an encouraging trend indicating a reduction in the gap between men and women. An exception would be Chemistry, where the gender gap increased over time. In Electrical, Electronic & Computer Engineering, the proportion of females funded has even surpassed that of males in recent years. Additionally, in Veterinary Science, male researchers had a higher funding ratio in the initial years, but this ratio has decreased recently, approaching that of female researchers. However, Figure 2 shows that while the average funding for males has increased in recent years, it was higher for females in earlier years. This suggests that despite female researchers receiving higher average funding, the proportion of female researchers funded has been lower compared to males.

Using a chi-square test to determine whether there is a significant difference between the funding ratios for each gender, this test is usually used to check the significance of the connection between two variables in a certain dependency relation [Dura and Drigă (2017)]. Specifically, the test is performed against the existence of differences in the funding ratio between male and female researchers. The result of the chi-squared test shows a p-value of 4.796e-107. Under a 5% significance level, this result provides strong evidence to reject the null hypothesis, indicating that there is a statistically significant difference in the funding ratio between male and female researchers.

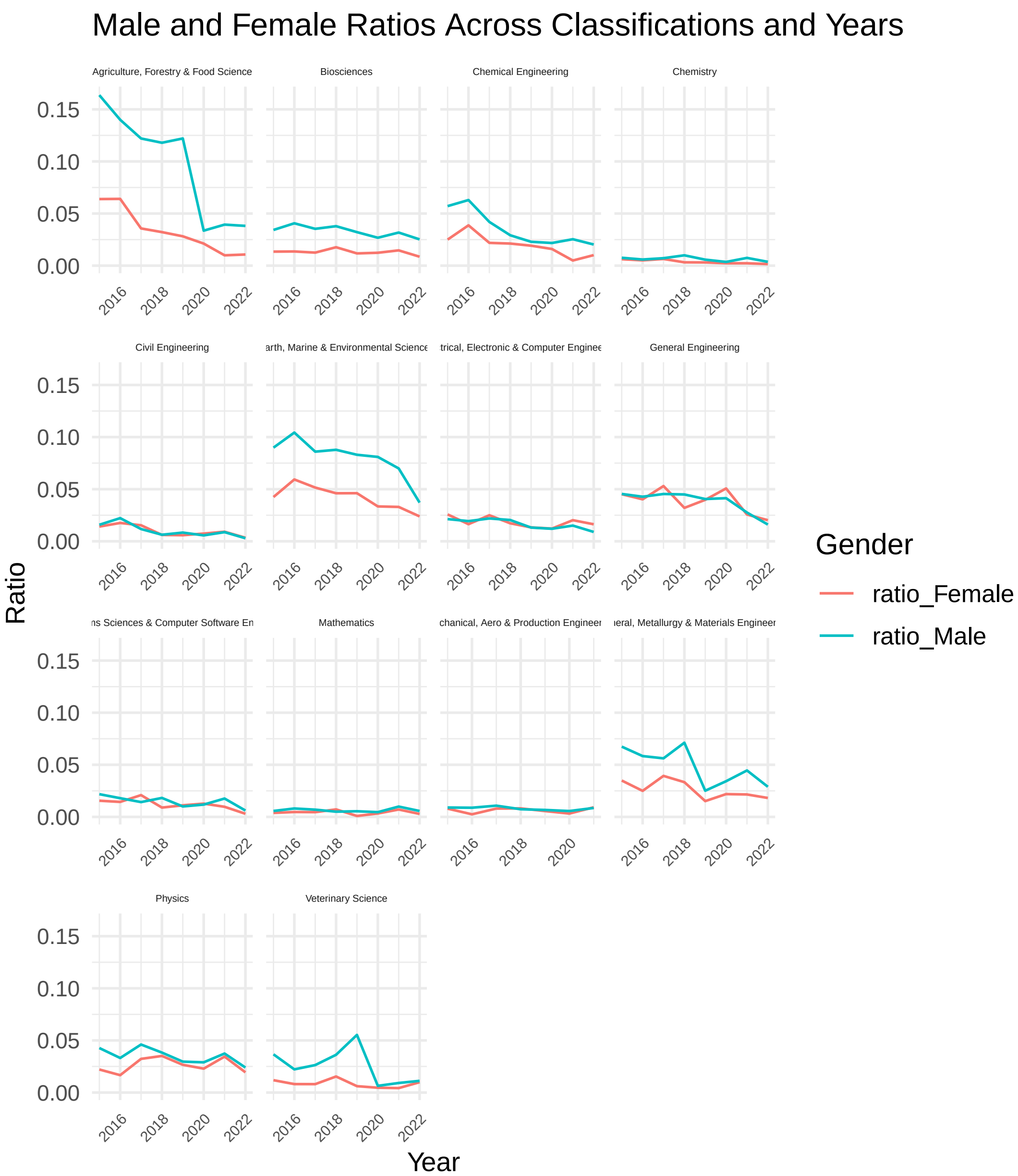

Figure 4: Trends in the proportion of each gender getting funded in across categories over time. In this result, the funding proportion is calculated by dividing the funded males or females by the total number of HESA staff in each gender.

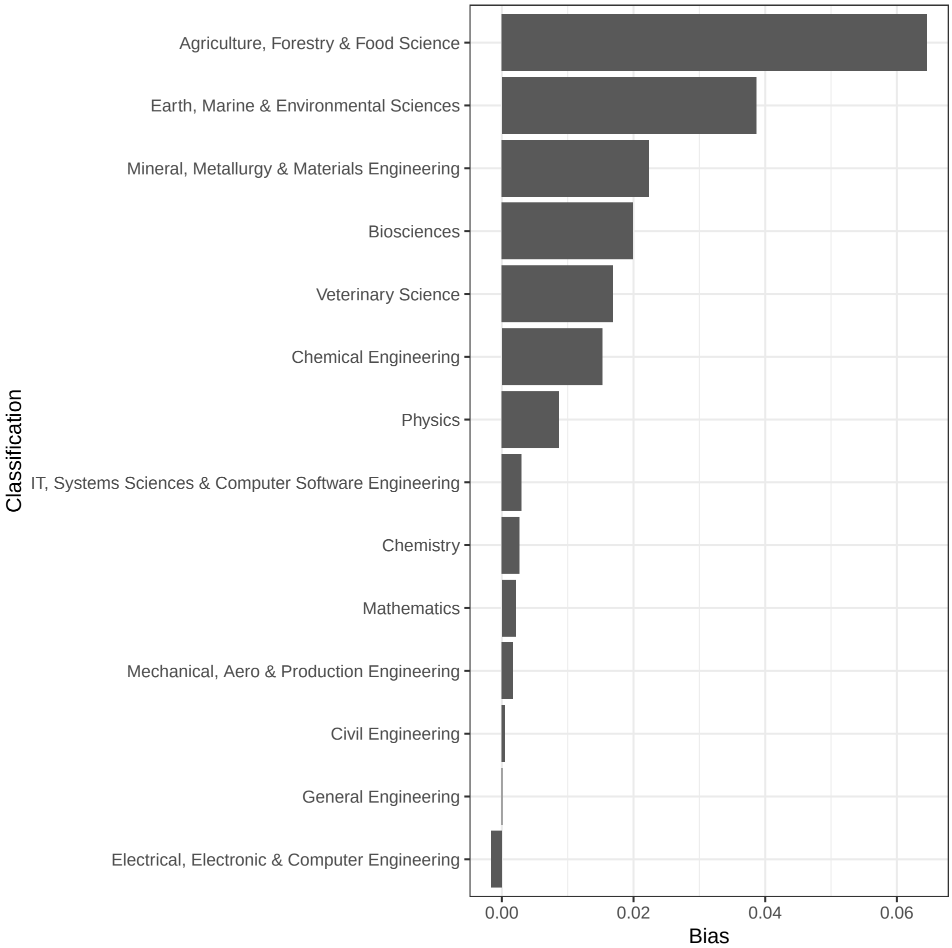

Figure 5: The ranking of the extent of gender bias in each classification over time. For the bias degree, I calculated the total number of funded males or females from 2015 to 2022, divided by the total number of HESA staff in the corresponding gender during the period. Then, the difference of the ratios for males and females is conducted and is reordered depending on the degree of biases - the upper represents the classification with greater gender bias.

4 Discussion

The findings of this study point to a gender disparity in the overall pattern of research funding within the UK STEM sector, no matter the aspect of funding amount or funding ratio. A positive aspect is the partial reduction of this gap, which could be attributed to the growing focus on funding diversity. However, the beneficial impact is not evident and inequities persist. On the UKRI's official website, a commitment to gender equality is outlined, detailing the equality plan from 2022 to 2026. This plan encompasses transparency and monitoring of gender data, support for female leaders in the UK, efforts to address gender inequality in recruitment, etc. [UKRI (2022)] However, to my knowledge, no actions in the plan specifically and directly address fair research funding between genders. Access to these resources remains unequal among different genders.

This study also compared the trend across STEM categories over the years. The results showed several observations: 1) Despite the general trend of reducing the gender gap, the gap in Chemistry increased in recent years; 2) The gender gap in the funding ratio in Biosciences persists; 3) Though most engineering disciplines have a gender gap in funding amount, the gap in funding ratio is small; 4) The gap in Veterinary Science shows an opposite pattern, where females receive higher average amount of funding but lower funding ratio compared to males.

The findings indicate a decline in the gap across most categories, with the exception of Chemistry, where the disparities in both funding amount and funding ratio in recent years have grown compared to 2015. According to a 2018 report by the Royal Society of Chemistry, the relative percentage of female chemists advancing from undergraduate studies to senior academic roles has decreased by 35 percentage points [RSC (2018)]. The primary causes identified include arbitrary funding criteria and a lack of transparency in the recruitment and promotion process [RSC (2018)]. Despite efforts to reduce the gender gap [RSC (2023)], the gap only slightly decreased and continued beyond 2018. Biosciences exhibit a similar trend; the persistent funding ratio gap reveals no notable increase in the gender gap. Positively, there is no apparent gender disparity in the funding amount for funded researchers. It was noted that female researchers in biosciences are generally less successful in obtaining research funding. Previous research has shown that women with biological sciences degrees are less likely than men to work as scientists post-degree and tend to have shorter publishing careers, primarily due to leaving the field, and lower publishing rates as well [Malespina and Singh (2023)]. Contributing factors may include: 1. Gender disparities in the recruitment and promotion of women in biological sciences, similar to chemistry; 2. Differences in life goals between males and females may influence their decision-making [Elliott (2016)].

Another key result from my analyses is the gender disparity in funding amounts for engineering disciplines. Unlike the situation in Biosciences, the gender gap here is mainly observable in the amounts awarded, while gender equality in the funding ratio is relatively higher. This suggests that the funding structure could be a key factor in the gender imbalance in this sector. Based on information from the UKRI Engineering and Physical Sciences Research Council (EPSRC) EPSRC (2023a), there are significant differences in the sizes of grants applied for by different genders, with female researchers consistently applying for smaller grants. Furthermore, the report points out that from 2007 to 2018, the award rates by value became increasingly different as the grant amounts increased EPSRC (2023a). The award rate for women remained consistently around 30% for all grant amounts, whereas for men it rose progressively from about 30% to nearly 60% as the grant amount increased. An EPSRC survey of female applicants revealed several reasons that might act as barriers to applying for larger grants EPSRC (2023b): 1. Lack of time due to being overwhelmed by other responsibilities; 2. Unfair biases in the peer review process; 3. Entry requirements for applications, such as institutional ranking or existing grant portfolio; 4. Absence of institutional support for large grant proposals. Notably, two of these primary causes are institution-related, which could indicate a troubling situation for female researchers. Previous research has shown that women are generally less likely to secure positions in highly ranked institutions compared to men Fox (2020). Consequently, even though higher-ranked institutions may offer benefits to their female staff, the barriers to entering these institutions might prevent many women from accessing these advantages. This highlights potential inequities in resource allocation and possible biases in admission criteria.

The funding landscape in Veterinary Science reveals contrasting patterns: The average amount of funding for female researchers is generally higher than for male researchers, particularly in the first three years; however, male researchers have a higher funding ratio. This may be related to the significant representation of women in veterinary science, as evidenced by data from the Royal College of Veterinary Surgeons, which reported that in 2021, 77% of practising veterinary surgeons in the UK were women [RCV (2022)]. Nevertheless, women are significantly underrepresented in journal publications - In 2023, 68% of authors in a main journal of veterinary science, namely veterinary surgery, were menBlacklock (2024)]. The gendered masculine culture and the barriers mentioned in the previous paragraphs could be potential reasons for this phenomenon [Liu et al. (2021)]. In addition to the comparatively higher interest in this field by women, veterinary science is often considered a caring profession. Previous literature has proved the gendered nature in care professions, with women being more prevalent in part due to gender stereotypes such as men being less compassionate but more professional than women [Poole and Isaacs (1997)]. This also sheds light on why fewer women pursue academic careers in Veterinary Science, despite their high participation and graduation rates in veterinary programs, and why female academics are more likely to hold lower-ranked positions compared to their male colleagues [Liu et al. (2021)].

4.1 Implications and Recommendations

In conclusion, our results revealed a significant gender bias in STEM research funding within the UK. Although there is a general decreasing trend for the gap, the academic framework continues to pose challenges to women. The internal link between research funding and the ranking of their institution could pose additional hurdles for female researchers. This could be regarded as a systematic bias in the review criteria, which benefits male applicants due to previous accumulative advantage [Witteman et al. (2019)]. This inequality not only causes a reduction in the female researcher's incentive to deliver innovative studies that contribute to science, but would also be a factor influencing the career life of female researchers [Jebsen et al. (2022)]. The number of research fundings, as an important role that affects the promotion of academic careers, would directly impact the retention and progression of female researchers [Jebsen et al. (2022)], forming a vicious circle harmful to the situation of female researchers. To break the cycle, a systematic improvement in gender equality is necessary, which could include establishing a diversity focus group that monitors the diversity of the employment of each institution and ensures that the university recruitment process is fair and offers the same opportunity to various groups of people [Sardelis et al. (2017)]. At the same time, this focus group should also be responsible for revising the funding application criteria, ensuring that the criteria could minimize the impact of accumulated advantage.

4.2 Limitations

In this study, we conducted the trend comparison using the ratio of each gender getting funded, where the HESA staff data is used for the calculation. We assumed that the HESA staff number in each classification and university is the total number of researchers in these areas. In addition, another caveat that is worth noticing is that this study only considered females and males as the genders. All the other genders are not listed in consideration, while these are also worth more attention. Though this study mainly focuses on female researchers, it can be a broad indication of the situation of most minority groups [Jebsen et al. (2022)]. Lastly, the gender determination for each project is based on the name of the primary applicant, which could introduce some inaccuracy.

4.3 Future directions

In this study, I have shown that there are gender biases among STEM disciplines, considering the total number of HESA staff to account for bias at all stages. However, to check the existing problem in the current diversity system, it would also be beneficial to study the funding trends based on the total number of applications so that we can find the proportion of bias caused during different stages, including the bias in the review stage and the step before application. In addition, this study only learned about the gender gap in STEM research funding. According to previous research, there may also be many other biases that may exist in research funding, including racial or nationality bias, ethnicity bias, risk aversion bias, etc. [Wojick and Michaels (2015)] All diversity should be as important as each other [Formanowicz et al. (2018)], and bias in any aspect would have a direct impact on science innovation and career life of individuals. As such, more studies on other elements of discrimination should be conducted to increase awareness.

5 Data and Code Availability

In my study, the HESA staff data is sourced from the official HESA website, including staff number under various institutions and categories over time. The project information from UKRI can be assessed from Onedrive. The coding part is available on Github, including the machine-learning related part and the analysis part.

References

Rcvs workforce summit 2021 - recruitment, retention and return in the veterinary profession. *RCVS*, 2022.

F. Bellotto Trigo, M. Mustri, R. Cornford, T. Clegg, S. Pawar, and W. Pearse. Placing uk research within the international stem funding landscape. Imperial College London, 2023.

K. Blacklock. Underrepresentation of a majority: Why are there so few women in senior roles in veterinary medicine? https://gender-test.sps.ed.ac.uk/blog/2024/underrepresentation-majority-women-veterinary-medicine-1, 2024.

J. H. Brown, J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West. Toward a metabolic theory of ecology. *Ecology*, 85(7):1771–1789, 2004.

A. Cislak, M. Formanowicz, and T. Saguy. Bias against research on gender bias. *Scientometrics*, 115: 189–200, 2018.

E. Commission. Gender equality in research and innovation. *European Comission*, 2020.

L. Craig and A. Powell. Non-standard work schedules, work-family balance and the gendered division of childcare. *Work, employment and society*, 25(2):274–291, 2011.

J. Crank. *The Mathematics of Diffusion*, volume 2. Clarendon press Oxford, 1975.

L. Cruz-Castro, D. K. Ginther, and L. Sanz-Menendez. Gender and underrepresented minority differences in research funding. 2022.

J. M. Delaney and P. J. Devereux. Understanding gender differences in stem: Evidence from college applications. *Economics of Education Review*, 72:219–238, 2019.

C. Dura and I. Drigă. Application of chi square test in marketing research. *Annals of the University of Petrosani. Economics*, 17(2):79–90, 2017.

S. L. Elliott. From the editor-in-chief: questions of gender equity in the undergraduate biology classroom, 2016.

EPSRC. Understanding our portfolio: A gender perspective. *UKRI*, 2023a.

EPSRC. Gender diversity in our portfolio: survey findings and interventions. *UKRI*, 2023b.

M. Formanowicz, A. Cislak, and T. Saguy. Research on gender bias receives less attention than research on other types of bias. *LSE Impact Blog*, 2018. URL https://blogs.lse.ac.uk/impactofsocialsciences/2018/05/17/research-on-gender-bias-receives-less-attention-than-research-on-other-types-of-bias/.

M. F. Fox. Gender, science, and academic rank: Key issues and approaches. *Quantitative Science Studies*, 1(3):1001–1006, 09 2020. ISSN 2641-3337. doi: 10.1162/qss_a_00057. URL https://doi.org/10.1162/qss_a_00057.

J. García-González, P. Forcén, and M. Jimenez-Sanchez. Men and women differ in their perception of gender bias in research institutions. *PloS one*, 14(12):e0225763, 2019.

A. García-Holgado, C. González, and A. Peixoto. Bridging the diversity gap in stem. In *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*, pages 193–195, 2019a.

A. García-Holgado, S. Verdugo-Castro, M. C. Sánchez-Gómez, and F. J. García-Peñalvo. Trends in studies developed in europe focused on the gender gap in stem. In *Proceedings of the XX international conference on human computer interaction*, pages 1–8, 2019b.

L. Hagen. Content analysis of e-petitions with topic modeling: How to train and evaluate lda models? *Information processing & management*, 54(6):1292–1307, 2018.

I. M. Handley, E. R. Brown, C. A. Moss-Racusin, and J. L. Smith. Quality of evidence revealing subtle gender biases in science is in the eye of the beholder. *Proceedings of the National Academy of Sciences*, 112(43):13201–13206, 2015.

M. Hosseini and S. Sharifzad. Gender disparity in publication records: a qualitative study of women researchers in computing and engineering. *Research Integrity and Peer Review*, 6:1–14, 2021.

J. Jebsen, K. Nicoll Baines, R. Oliver, and I. Jayasinghe. Dismantling barriers faced by women in stem. *Nature Chemistry*, 14(11):1203–1206, 2022.

J. M. Jebsen, C. Abbott, R. Oliver, E. Ochu, I. Jayasinghe, and C. Gauchotte-Lindsay. A review of barriers women face in research funding processes in the uk. *Psychology of Women and Equalities Review*, 3(1-2), 2020.

J. Kang, J. Hense, A. Scheersoi, and T. Keinonen. Gender study on the relationships between science interest and future career perspectives. *International Journal of Science Education*, 41(1):80–101, 2019.

S. Kong, K. Carroll, D. Lundberg, P. Omura, and B. Lepe. Reducing gender bias in stem. *MIT Science Policy Review*, 1(8):55–63, 2020.

D. Kube, J. Weidlich, K. Kreijns, and H. Drachsler. Addressing gender in stem classrooms: The impact of gender bias on women scientists' experiences in higher education careers in germany. *Education and Information Technologies*, pages 1–28, 2024.

Y. Lee, R. M. Capraro, and A. Bicer. Gender difference on spatial visualization by college students' major types as stem and non-stem: a meta-analysis. *International journal of mathematical education in science and technology*, 50(8):1241–1255, 2019.

X. Liu, R. Dunlop, R. Allavena, and C. Palmieri. Women representation and gender equality in different academic levels in veterinary science. *Veterinary Sciences*, 8(8):159, 2021.

E. López-Iñesta, C. Botella, S. Rueda, A. Forte, and P. Marzal. Towards breaking the gender gap in science, technology, engineering and mathematics. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, 15(3):233–241, 2020. doi: 10.1109/RITA.2020.3008114.

A. Malespina and C. Singh. Gender gaps in grades versus grade penalties: Why grade anomalies may be more detrimental for women aspiring for careers in biological sciences. *International Journal of STEM Education*, 10(1):13, 2023.

C. A. Moss-Racusin, C. Sanzari, N. Caluori, and H. Rabasco. Gender bias produces gender gaps in stem engagement. *Sex Roles*, 79(11):651–670, 2018. doi: 10.1007/s11199-018-0902-z. URL https://doi.org/10.1007/s11199-018-0902-z.

M. Poole and D. Isaacs. Caring: A gendered concept. In *Women's Studies International Forum*, volume 20, pages 529–536. Elsevier, 1997.

M. Ranga, N. Gupta, and H. Etzkowitz. Gender effects in research funding. *Bonn: Deutsche Forschungsgemeinschaft*, 2012.

RSC. Breaking the barriers: Women's retention and progression in the chemical sciences. https://www.rsc.org/globalassets/02-about-us/our-strategy/inclusion-diversity/womens-progression/media-pack/v18_vo_inclusion-and-diversity-_womans-progression_report-web-.pdf, 2018. Accessed: 10 Aug 2023.

RSC. Framework for action to improve inclusion and diversity in scientific publishing, 2023. URL https://www.rsc.org/policy-evidence-campaigns/inclusion-diversity/framework-for-action/#framework. Accessed: [10 Aug 2023].

S. Sardelis, S. Oester, and M. Liboiron. Ten strategies to reduce gender inequality at scientific conferences. *Frontiers in Marine Science*, 4:231, 2017.

M. Thelwall, M. Abdoli, A. Lebiedziewicz, and C. Bailey. Gender disparities in uk research publishing: Differences between fields, methods and topics. *Profesional de la información/Information Professional*, 29(4), 2020.

UK Space Agency. Government announces plans for largest-ever r&d budget. https://www.gov.uk/government/news/government-announces-plans-for-largest-ever-rd-budget, 2024.

UKRI. Gender equality plan 2022 to 2026. https://www.ukri.org/publications/ukri-gender-equality-plan/gender-equality-plan-2022-to-2026/, 2022. Accessed: 10 Aug 2023.

S. Verdugo-Castro, A. García-Holgado, and M. C. Sánchez-Gómez. The gender gap in higher stem studies: A systematic literature review. *Heliyon*, 8(8), 2022.

M.-T. Wang and J. L. Degol. Gender gap in science, technology, engineering, and mathematics (stem): Current knowledge, implications for practice, policy, and future directions. *Educational psychology review*, 29:119–140, 2017.

P. White and E. Smith. From subject choice to career path: Female stem graduates in the uk labour market. *Oxford Review of Education*, 48(6):693–709, 2022.

Wikipedia. Latent dirichlet allocation, 2023. URL https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation. Accessed: [Your Access Date Here].

WIPO. Global innovation index 2022 - what is the future of innovation-driven growth? *WIPO*, 2022.

H. O. Witteman, M. Hendricks, S. Straus, and C. Tannenbaum. Are gender gaps due to evaluations of the applicant or the science? a natural experiment at a national funding agency. *The Lancet*, 393 (10171):531–540, 2019.

D. E. Wojick and P. J. Michaels. Is the government buying science or support? a framework analysis of federal funding-induced biases. *Washington DC, Cato Institute, Cato Working Paper*, (29), 2015.

S. Women. Women in stem statistics: Progress and challenges. *STEM Women*, 2023.