# Introduction

Gender discrimination is a problem, especially in the technology industry. A recent study by [CITE] looked at GitHub, one of the most popular software repositories with 14 million users, to see if the effect can be seen by comparing pull requests between male and female contributors. A pull request happens when someone proposes changes on a software code repository that is being hosted on GitHub. The request can either be accepted or denied by the owners of the repository. The researchers found that the approval rates of pull requests, when made without revealing one's gender, were comparable between both genders. The rate of approval for women fell significantly if the coder's gender was identifiable upon requesting a pull whereas the rates for men did not change much.

Wikipedia, the largest encyclopedia mankind has ever created, also has a gender problem. Estimations about exact numbers are difficult to find but it is undisputed, that most contributors and even editors are of male gender. Especially the latter is important since editors, specifically veteran editors, do have the powers to reject articles and restrict access of other contributors which has a severe impact on the acceptance rate regarding articles that are written about female figures. The resulting bias with respect to the variety of content is increasingly recognized leading to Wikipedia's gender gap to be an ever-increasing topic of interest, drawing the attention of scholars of new media and similar fields.

Discussing the systemic bias on Wikipedia is important as the platform poses the most powerful source of information worldwide. It is available in 275 languages and is being accessed by millions of people regularly. Many assistive devices and applications such as Smartwatches, Google Assistant, Alexa, Siri or Cortana draw their knowledge directly from Wikipedia without informing their users about the source or authors. Through these developments, the world, as it is represented by Wikipedia, is widely being considered as natural.

In this work, we address gender bias discrepancies within the technology industry by providing facts on historical developments and current events. Additionally, we try to measure possible trends for the future of digital environments that have collaborative characteristics, draw implications for the quality of publicly available information and participation rates and provide suggestions that may have an impact on future developments.

# Considerations and Limitations

Most studies on gender bias have been conducted either in Europe or in the United States. Only very few researchers actively work on that topic from other parts of the world. The participants of the studies conducted are more likely to be people from those geographical regions. Thus, their provided data and insights inherently carry strong biases and are not globally representative. A lot of studies in this area are user studies, which poses an issue in our specific case as it can be shown [CITE], that women are overall less likely to participate in user studies compared to men. As an example, this phenomenon has an impact on the data used in Figure 1 since the underlying studies could be skewed accordingly.

# Wikipedia

In the following we examine several issues connected to Wikipedia's gender gap and different reasons that are responsible for the low contribution rates of female participants or the low

acceptance rate regarding articles about women. We then also make suggestions on how to combat these issues effectively.
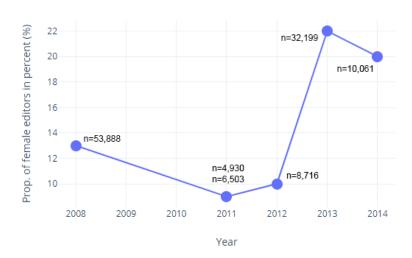


*Figure 1: 6 Studies were conducted by the Wikimedia Foundation between 2008 and 2014 to measure the proportions between male and female editors that are active on Wikipedia. The corresponding number of participants is annotated for each study accordingly.*

## Low proportions of female editors

The Wikimedia Foundation conducts various user studies at annual intervals, and the data thus collected is made publicly available. Part of these studies is also to determine the ratio of female to male editors on Wikipedia. We analyzed the results of those studies to see how the ratios have changed over time. Our results can be seen in Figure 1. We found that women are strongly underrepresented on Wikipedia, even in the strongest estimate with 22% of editors being women in 2013. However, a positive trend can also be observed from the year 2011 onward.

## Self-perception as a major cause for the gender gap

A characteristic with a particularly strong influence on the interest in participating in the processing of publicly accessible information is self-perception. Helgeson et al. [CITE] showed, that there are significant differences in self-perception and confidence between both genders. The reason for that is unknown. This difference could be due to women being socialized to be less self-confident or that their experiences lead them to become less self-confident, as the authors note, but this needs to be subject for further studies and investigations as the paper did not aim to find the cause for the discrepancy found.

Hinnosaar et al. [CITE] showed, that the gender gap in Wikipedia editing is, to a large share, because one's belief about competence and draws parallels to the findings of Helgeson et al. [CITE]. These findings are supported by the studies of Collier & Bear et al. [CITE] and Protonotarios & Sarimpei et al. [CITE], but it cannot be shown if the reasons for that lie within the internal structures of Wikipedia, such as sexism within the Wikipedia community, or if these are caused by external factors such as societal standards and expectations.

## The influence of conflicts on female contributions

Lam & Uduwage [CITE] performed a variety of quantitative analyses on publicly available English Wikipedia articles and found, that new users, who could be publicly identified as female, were more likely to stop editing and leave Wikipedia when their edits were reverted. This poses an issue regarding the balance of equal representatives on Wikipedia since the number of female editors

attempting to join Wikipedia is already very low. They also found that articles, where female editors dominated on the amount of contributions, had significantly more disagreements in discussions around edits than other articles [CITE SAME AS BEFORE]. Being subject to large amounts of conflicts causes discomfort and demotivation when it comes to editing other people's work, since the editors fear to receive even more critical feedback and have their contributions being reverted.

## Article rejection due to notability criteria

Wikipedia consists of multiple categories for different types of topics. Although all articles seem to share a similar styling, they belong to different categories. A user can navigate through those categories and see lists of articles belonging to any selected category. For example, there is a category on computer scientists, where a user can find various related lists of scientists such as lists by nationality. Such categories often have their own communities of editors that are active in the corresponding field who can sometimes impose very harsh acceptance criteria, so that the attempt to publish articles in the respective area often gets rejected due to the lack of fulfillment of said criteria. Wagner et al. [CITE] found, that there are many areas in Wikipedia where the criteria are unfavorable towards female figures. This is the reason why many articles about women are rejected which further leads to discrepancies when it comes to the ratio of male to female representation regarding Wikipedia articles.

But Klein et al. [CITE] shows that this situation is getting considerably better as societal changes take place with the feminism movement gaining more traction and with public attention being drawn towards a more considerate behavior towards other genders. Initiatives such as meet-ups, talks and competitions with special emphasis on female participants show a net positive development when it comes to public perception and awareness towards females and women's rights worldwide fostering opportunities for women to overcome societal hurdles and gain equal access to positions within higher ranking institutions and well paid job opportunities. Such societal changes are strongly reflected within Wikipedia as well.

## Content differences between articles about men and women

Wikipedia allows to link to other articles in the text of an article. This simplifies explanations of terms and connects coherent, relevant information without packing everything into one page and losing the overview. Wagner et al. [CITE] found, that these internal links within articles about women frequently lead to articles about men while in comparison, men who are interlinked to women very often lack such internal Wikipedia links to the articles of the corresponding women. This makes it look as if it is women who are usually related to men and not the other way around. Relating to this issue, Graells-Garrido et al. [CITE] found, that articles about women include the *spouse* attribute with reference to the person significantly more often than articles about men do. Recommendation and search algorithms depend on strong connections between Wikipedia articles and the lack of those can lead to women being discriminated when it comes to ranking articles regarding their notability.

Even the language used within articles compared between men and women is different. Wagner et. al and Graells-Garrido et al. both showed that articles about women contain significantly more information about relationship and family issues than articles about men do. This may be due to societal issues such as women often being related to the home domain and relationship domain.

## How to combat the gender gap on Wikipedia

To prevent newcomers from being further scared off because of conflicts, Wikipedia editors must be more careful with their power to discard proposed changes. This aspect should be included in the Wikipedia guidelines and clearly emphasized. Proposed changes of positive intent should be encouraged by giving more thought to thanking users for their contributions. Such a reaction is also publicly visible and of great interpersonal importance. It can also strengthen the self-esteem of the

users concerned, thereby improving the perception of the competence of newcomers. This in turn could be beneficial in terms of motivation to edit existing articles or to create new articles.

To further battle the gap, increasing efforts need to be put to invite women as contributors on Wikipedia. This can be done through explicit information events that focus on women. There are independent associations around the world that are affiliated with the Wikimedia Foundation and promote Wikipedia's mission to facilitate the distribution of freely available information. One such organization exists in Vienna, for example, and stages weekly open access information events where anyone can visit their office and learn how to write or edit articles on Wikipedia. Such associations need to put more effort into attracting the attention of women and encouraging their contributions, considering the use of direct and indirect marketing methods with appropriately targeted content. In addition, opportunities could be created for women to socialize with each other.

When it comes to notability criteria, editors need to double check their stance on the importance of such figures. It often happens that women who have made meaningful achievements within their communities and therefore are of great value to recognized groups of people do not receive equal recognition from the Wikipedia community, even though their achievements are often recognized as being important and can be verified by trustworthy primary and secondary literature and other sources. This needs to be discussed and considered by editors of different sections and categories within Wikipedia.

Articles about women, especially historic figures, need to be revised and added. As an example, Wikipedia lacks a lot of data about women from the 1940s to 1950s and existing articles suffer a lot from historic biases since the sources used were already biased. This requires serious investigative efforts to obtaining the facts and curate the information into Wikipedia.

# Language

The social influence of Wikipedia is undisputed. It is important to help raise public awareness and make people sensitive to the various forms of discrimination. This includes discrimination through language, including that used to write or edit articles on Wikipedia. Stereotypes and prejudices are made about women in all areas of life and social environments. This discriminatory use of language takes place in everyday conversations, in the media, in textbooks, on the labour market, etc. and is accepted, used and passed on, consciously or unconsciously, reflected or unreflected. Correspondingly stereotypical texts also exist in Wikipedia. The starting point for any discrimination is prevailing social norms, which are set by the majority of a society and do not take into account differences. Non-discriminatory language must be perceived and practised individually in everyday life. Language is often understood as a neutral means of communication in everyday life, but it is a very powerful tool when it comes to giving meaning and significance to our world. But it is a very powerful tool when it comes to giving meaning and significance to our world. We discriminate against others not only by what we do, but also by what we say and what we don't say. Wikipedia is particularly subject to implicit language discrimination, for example when relevant information is systematically not mentioned or irrelevant, stereotypical information is highlighted. Some common spelling rules should be taken into account when writing or editing articles. As an example: women in politics are often described as female politicians, whereas articles about male politicians do not mention gender. This gives the impression that male is the standard gender and that women are "something different" in comparison. When categorising articles, equal treatment must be ensured, because a similar effect can be caused by, for example, moving female politicians into a separate category called "female politicians" while leaving male politicians in the "politicians" category. This creates the impression that female politicians need an explicit adjective to distinguish them from

politicians in general. This is especially important in the lead part of the article, since most people, as well as software applications, pay most attention to that section. Likewise, one should avoid using expressions like "She was the first woman to ..." as they put gender above actual achievement and indirectly emphasize that there have been men who have already achieved the same, which could be perceived as reducing the value of the achievement. These formulations can have the effect that the respective woman has succeeded in achieving a less valuable form of said achievement, but is still worthy of attention because of her femininity. Also, family references should be used just as sparingly with women as with men. The importance of a woman should not be defined exclusively in terms of her relationship with other men. Accordingly, care should be taken to ensure that mentions such as relationship status, marriage, divorce or sexuality are mentioned in a meaningful, context-related way. If it is considered appropriate to mention marriage to a man, then expressions such as "man and wife" should be avoided. Instead of "A is the wife of B", "A is married to B" should be used as a form of expression, since otherwise the man is treated in a generalizing manner, but the woman is explicitly marked.