

Mitigating Gender Bias in Search Engines

Vishal Goud Mogili
vmogil2@uic.edu
University of Illinois at Chicago
Chicago, Illinois, USA

Hemalatha Ningappa
Kondakundi
hning4@uic.edu
University of Illinois at Chicago
Chicago, Illinois, USA

Niketan Doddamani
ndodd@uic.edu
University of Illinois at Chicago
Chicago, Illinois, USA

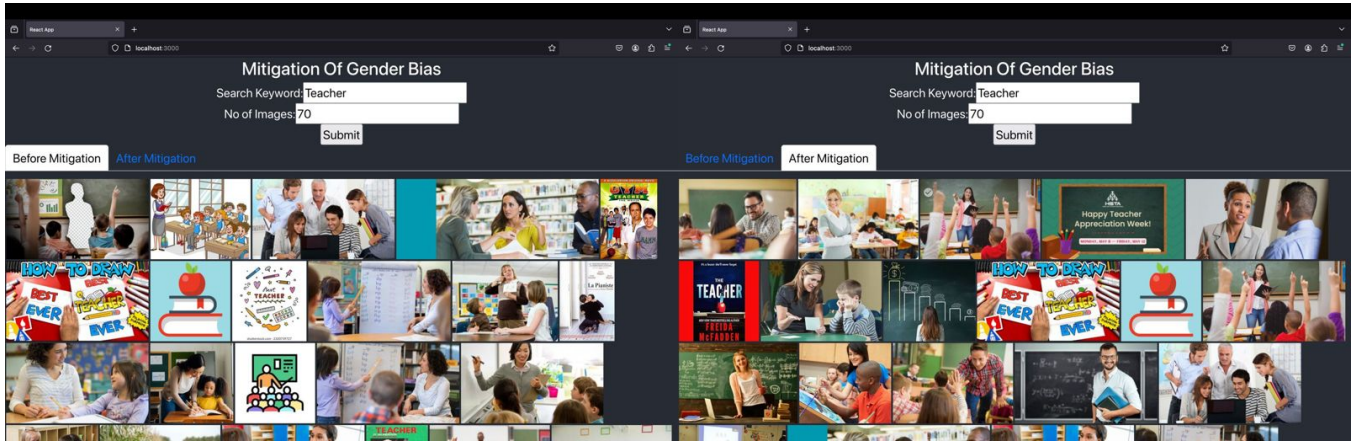


Figure 1: Custom User Interface for Image Search.

ABSTRACT

This project investigates the mitigation of gender bias in automated image labeling systems, building upon previous research. While their study utilized Amazon Rekognition to assess gender representation in image search results, our project employs a modified version of the ResNet-10 architecture for the detection task. Our research focuses on the examination and mitigation of gender bias across four major image search engines like Google, Baidu, Naver and Yandex. In addressing the limitations of existing gender detection APIs, we used a modified version of ResNet-10 architecture pre-trained neural network model for gender classification. Furthermore, we design and implement three novel re-ranking algorithms—the Epsilon-Greedy Algorithm, the Relevance-Aware Swapping Algorithm, and the Fairness-Greedy Algorithm to address the gender bias evident in the visual representation of professional occupations. Our findings demonstrate the persistent and un-systematically addressed nature of gender bias in image search engines, underscoring the need for a systematic and sustainable approach to bias mitigation. The project identifies the importance of non-binary gender recognition and the need for inclusive AI systems, highlighting the urgency for continued research in this domain.

KEYWORDS

Gender Bias, Search Engines, Image Retrieval, Re-ranking Algorithm, Diversity, Inclusion.

1 INTRODUCTION

In the digital age, image search engines are a primary avenue for individuals to access visual information. The images returned by these searches are not mere reflections of queries but also shape societal perceptions, especially regarding gender roles and representations [12]. With the advent of AI-driven image search engines, the potential for these systems to perpetuate and amplify gender stereotypes has become a significant concern. This project report builds upon the previous work of [5], who explored gender bias in image search results and proposed adversarial attack queries to investigate the depth of the issue. Our study extends this line of inquiry by employing a modified version of the ResNet-10 architecture [19], in contrast to [5] use of the Amazon Rekognition API. By analyzing image search results through the lens of a different AI model, we aim to provide a more comprehensive understanding of how gender representations are constructed and disseminated by automated systems. The ResNet-10 architecture, known for its deep learning capabilities in image recognition tasks, has been adapted in our project to serve as a tool for identifying gender within image searches. We acknowledge that images are powerful carriers of information, capable of reinforcing societal norms and shaping collective worldviews [13]. When image search results present a skewed gender distribution, especially in professional contexts [10], they can subtly influence individuals' perceptions about gender roles. Recognizing the role that image search engines play in either reinforcing or challenging gender biases, our study set out to systematically examine the representation of gender in professional occupations [18] through images returned by widely used search engines such as Google, Baidu, Naver, and Yandex. This

large-scale analysis allowed us to discern patterns and inconsistencies in how genders are depicted and to explore the efficacy of search engines in providing equitable gender representation. To improve upon the gender detection methodologies [20] employed in previous research [17], we adopted a hybrid approach. This method combines the efficiency of automated gender detection with the ground truth through public census data. Our approach aims to address the shortcomings of relying solely on AI, which can often overlook the nuances of gender expression and identity. In addition to detection, our project emphasizes the need for mitigation strategies to counteract gender bias. We propose and evaluate three novel re-ranking algorithms designed to redress the identified biases. These algorithms not only serve to re-rank search results to present a more balanced gender representation but also highlight the importance of developing systematic approaches for long-term solutions. Our investigation into gender bias in AI-driven image labeling systems offers insights into the current state of bias mitigation. By critically assessing and proposing solutions to these biases, our work contributes to the broader discourse on ethical AI practices and the pursuit of fairness in digital information spaces. Through this report, we hope to inform and inspire further research and development efforts that aim for a more inclusive and unbiased portrayal of genders in AI applications.

2 PROBLEM STATEMENT

In the realm of image search engines, persistent gender biases often manifest in search results, even when gender-neutral queries are employed. Despite efforts by search engine providers to rectify these biases, they continue to resurface, influencing societal perceptions and reinforcing stereotypes. This recurring issue undermines the objective of providing equitable and inclusive representations in search results. Traditional approaches [20] to *Mitigating gender bias* in image search engines have proven insufficient and lack universality across different search contexts. Existing solutions are often short-lived and fail to address bias comprehensively, leading to its continued prevalence. Furthermore, the effectiveness of bias mitigation efforts varies significantly depending on the search engine utilized and the specific search terms employed, indicating a need for more robust and adaptable strategies. Therefore, there is an urgent need for innovative solutions that can accurately detect and effectively mitigate gender biases in image search results across diverse search contexts. These solutions must be capable of providing fair and inclusive representations while addressing the inherent challenges posed by variations in search engine algorithms and user query patterns. In summary, the problem statement revolves around the persistent presence of gender biases in image search results, despite attempts to mitigate them. Addressing this issue requires the development of novel techniques that can adapt to different search contexts and provide equitable representations across diverse user demographics.

3 METHODOLOGY

In the following sections, we will delve into the specifics of each component, detailing the technologies, frameworks, and algorithms employed to realize our objectives. From the user interface design

using ReactJS to the backend logic implemented in Flask, each aspect of our technical implementation is meticulously crafted to ensure seamless functionality and optimal performance. Additionally, our approach leverages pre-trained neural network models for image processing tasks and incorporates state-of-the-art bias detection and mitigation algorithms to address gender biases effectively.

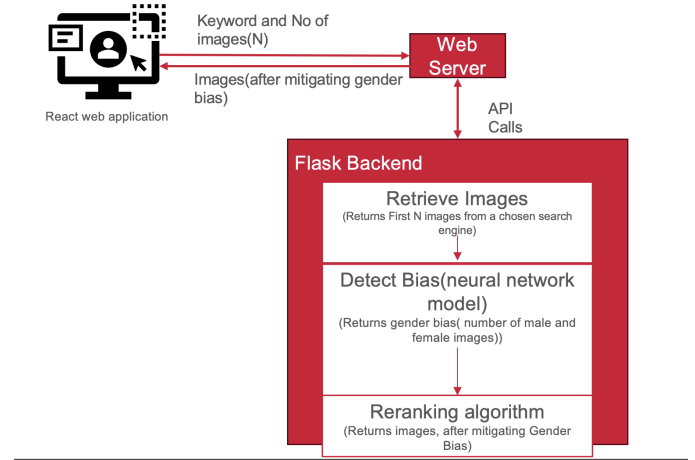


Figure 2: Architecture Diagram.

3.1 Fixing The Ground Truth Value:

Algorithm 1 Adjustment of Gender Representation Ratios

- 1: Initialize ground truth ratios:
- 2: $male_ratio \leftarrow 0.5, female_ratio \leftarrow 0.5$
- 3: Total number of recognized samples: $total_samples$
- 4: Number of male positive samples: $male_positives$
- 5: Number of female positive samples: $female_positives$
- 6: Observe the trend in the first 10 samples
- 7: **if** $male_ratio < female_ratio$ **then**
- 8: $male_ratio \leftarrow male_ratio \times 1.20$
- 9: $female_ratio \leftarrow female_ratio \times 0.80$
- 10: **else**
- 11: $female_ratio \leftarrow female_ratio \times 1.20$
- 12: $male_ratio \leftarrow male_ratio \times 0.80$
- 13: **end if**
- 14: Set the new ground truth using updated $male_ratio$ and $female_ratio$
- 15: Rerank the images based on the new ground truth
- 16: Continue the process until the first 10 images show balanced representation without bias

3.2 The Front-End User Interface (React Web Application):

The user interface of our image search application (figure:1) is developed using ReactJS[4], a popular JavaScript library known for its

flexibility and responsiveness. By leveraging React, we ensure that our application offers a smooth and interactive experience to users across different devices and screen sizes. The design incorporates an intuitive input form where users can specify their search terms and the number of images they wish to retrieve. This input is captured by the front-end module, which then sends the requests to the server using Fetch API calls. Upon receiving the processed results from the server, the user interface dynamically updates to display the mitigated images, providing users with real-time feedback on their search queries.

3.3 Server-Side Processing (Web Server):

At the heart of our image search application lies the web server[4], which serves as a robust API gateway for handling client requests and orchestrating the various backend services. Built with a focus on RESTful principles, the server efficiently routes incoming requests from the front end to the appropriate backend components. It handles seamless communication between the client-side application and the Flask backend. This modular design not only enhances the scalability and maintainability of our system but also lays the foundation for future enhancements and integrations with additional services or APIs.

3.4 Flask Backend:

Our backend logic is implemented using Flask[14], a lightweight and efficient web framework known for its simplicity and versatility. Flask provides a solid foundation for building web services, allowing us to focus on implementing the core functionalities of our image retrieval and processing pipeline. Leveraging Flask's extensibility, we have designed our backend to accommodate future enhancements and adaptations, such as integrating additional bias detection models or incorporating new search engine APIs. This flexibility ensures that our system remains adaptable to evolving requirements and technological advancements in the field of image search and bias mitigation.

3.5 Image Processing and Bias Mitigation

3.5.1 Retrieve Images: To retrieve images from various search engines, our application integrates with Google search APIs [7], utilizing API calls to fetch images based on user-specified search terms. The number of images retrieved is determined by the user's input, ensuring flexibility and customization in search results. We employ a strategic approach to image selection, incorporating filters for image type, resolution, and relevance provided by the search engine's API. This ensures that the retrieved images are diverse and representative, facilitating comprehensive analysis and bias detection.

Data collection involved an extensive process where queries were constructed to include a range of professional occupations. These queries were then input into several popular image search engines—Google[7], Baidu[1], Naver[2], and Yandex[3]. Each query's first 200 image results were collected. This diverse dataset ensures a comprehensive analysis across cultures and search algorithms, providing a global perspective on gender representation.

3.5.2 Detect Bias (Neural Network Model): For bias detection, we leverage a convolutional neural network (CNN) based on a

modified ResNet-10 architecture. The specific model we use is the "res10_300x300_ssd_iter_140000_fp16.caffemodel", which is pre-trained for face detection tasks [9]. This model is widely recognized for its accuracy and efficiency in detecting faces within images. It utilizes a modified version of the ResNet-10 architecture[8], incorporating residual blocks to handle deep networks effectively. Additionally, the model is trained using the Single Shot MultiBox Detector (SSD) framework[11], enabling it to detect objects directly from feature maps obtained at different scales without the need for a separate region proposal network. The iteration number 140000 indicates the number of training steps the model has undergone, ensuring robust performance in face detection tasks. Furthermore, the model utilizes 16-bit floating-point precision (FP16), reducing memory demand and computational requirements, making it suitable for real-time applications and devices with limited computational power.

3.5.3 Implementation of Re-ranking Algorithms: The working principles of the three re-ranking algorithms developed for mitigating gender bias in image search results are based on different core concepts, each tackling the problem of bias from unique angles.

(1) **Epsilon-Greedy Algorithm:** The Epsilon-Greedy algorithm [6] is a strategy derived from the domain of reinforcement learning, where the selection between exploring a new possibility or exploiting a known reward is randomized. In the context of our image re-ranking:

- **Exploration:** Introducing randomness in the ranking to disrupt gender biases and explore new gender distributions.
- **Exploitation:** Retaining the original order to exploit the relevance of the search results.

The algorithm assigns each image a probability of being swapped based on the epsilon value ϵ , which is a user-defined parameter that dictates the degree of randomness. For instance, with a higher ϵ , there is a greater chance that any given image will be swapped with another, thereby increasing the diversity of the output. The algorithm proceeds by iterating over the list of images and, for each one, deciding based on ϵ whether to swap it with another image further down the list.

(2) **Relevance-Aware Swapping Algorithm:** This algorithm is more sophisticated than the epsilon-greedy approach as it accounts for the relevance of images. The underlying idea is that while randomness can introduce fairness, it should not come at the cost of relevance:

- **Relevance Weight Modeling:** It assigns a weight [16] to each image based on its original rank, with the assumption that search engines display the most relevant images first. This weight influences the likelihood of the image being moved; images with higher relevance are less likely to be swapped.
- **Swapping Probability:** The algorithm generates a swapping probability for each image by considering both its relevance weight and a sensitivity parameter ρ . This parameter adjusts the balance between maintaining relevance and increasing randomness for fairness.

During execution, the algorithm iterates over the list, and for each image, it calculates the swapping probability and decides whether to swap this image with another, chosen at random from lower down in the list. The chance of swapping decreases as the relevance of an image increases.

- (3) **Fairness-Greedy Algorithm:** The fairness-greedy algorithm operates under a different paradigm, explicitly focusing on the fairness of the image ranking:
- **Fairness Objective:** It aims to achieve a ranking that mirrors the real-world gender distribution [15] of the occupation in question. This requires prior knowledge of the gender distribution (the ground truth), which can be obtained from sources such as census data.
 - **Re-ranking Process:** The algorithm examines the top-ranked images and compares their gender distribution with the ground truth. If discrepancies are found, the algorithm identifies images lower in the ranking that would correct this imbalance and promotes them within the list.

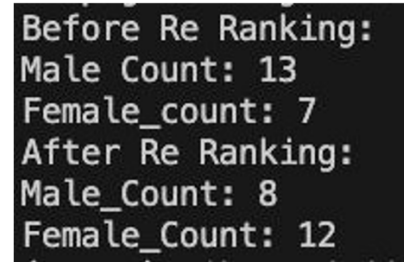
This process is conducted iteratively, starting from the top of the list and working downwards, ensuring that the most viewed images—those at the top of the search results—are as representative as possible.

Each of these algorithms was subjected to a series of evaluations using both synthetically generated data with known biases and real-world data from the image searches. The algorithms' effectiveness was measured by how well the re-ranked image lists matched the expected gender distributions. This evaluation was quantified using statistical measures, such as the normalized difference, to understand the degree of bias mitigation each algorithm could achieve. Through this detailed analysis, the efficacy and potential application of each algorithm in reducing gender bias in image search results were thoroughly assessed. Our methodologies and algorithms represent a systematic approach to understanding and mitigating gender bias in image search results. By leveraging a modified version of ResNet-10 and detecting genders in the images, our project provides an alternative viewpoint to existing studies. The re-ranking algorithms are crafted not just to randomize or rebalance search results but also to incorporate an understanding of fairness that aligns with societal expectations and professional gender distributions. This unique approach to detecting and mitigating bias could serve as a model for future efforts in creating more equitable AI systems.

3.5.4 Data Flow. The data flow within our system follows a structured sequence of operations, starting from the user input captured by the React web application and culminating in the display of reranked images. A detailed flowchart figure:2 illustrates this sequence, highlighting the asynchronous nature of the system and emphasizing the non-blocking I/O operations that enhance user experience by allowing the system to process in the background while the user interface remains responsive.

3.6 Additional Considerations

In addition to the core functionalities of our image search application, we have incorporated several additional considerations to enhance security, error handling, and performance. Robust security



```
Before Re Ranking:
Male Count: 13
Female_count: 7
After Re Ranking:
Male_Count: 8
Female_Count: 12
```

Figure 3: Epsilon-greedy algorithm. Introduces randomness to break clusters of biased results. (For example: top 20 male CEO image search results).

measures, such as HTTPS communication and input validation, protect against potential security vulnerabilities and injection attacks. Exception handling mechanisms ensure graceful management of interruptions in the image retrieval or processing pipeline, providing informative feedback to users in case of errors. Continuous performance monitoring and usage analytics enable us to identify and address bottlenecks, optimizing system performance and user experience over time.

4 RESULTS

This section presents the results obtained from the implementation of three distinct re-ranking algorithms aimed at mitigating gender bias in first 20 image search results. Each algorithm adopts a unique approach to ensure a balanced gender representation. Below we discuss the results before and after applying each re-ranking algorithm:

Epsilon-Greedy Algorithm: Initial Results: Before re-ranking, the search results showed a significant gender imbalance with 13 male and 7 female images.[5]

Post-Algorithm Results: After applying the epsilon-greedy algorithm, which introduces randomness to disrupt clusters of biased results figure:3, we observed a notable improvement in gender balance. The male count was reduced to 8, and the female count increased to 12.

Analysis: The epsilon-greedy algorithm's randomness effectively breaks up the concentration of biased results (such as male-dominated CEO images) and produces a more equitable gender distribution.

Relevance-Aware Swapping Algorithm: Initial Results: The gender distribution of search results before re-ranking consisted of 9 male and 11 female images.

Post-Algorithm Results: The application of the relevance-aware swapping algorithm yielded a modest improvement, resulting figure:5 in 8 male and 12 female images .

Analysis: This algorithm maintains a balance between relevance and randomness. The slight alteration in gender distribution post-re-ranking indicates that while maintaining the relevance of search results, the algorithm successfully addresses gender bias without compromising the integrity of the results.

Fairness-Greedy Algorithm: Initial Results: Initially, the results showed a disparity with 11 male and 9 female images.


```

Before Re Ranking:
Male Count: 9
Female_count: 11
After Re Ranking:
Male_Count: 8
Female_Count: 12

```

Figure 4: Relevance-aware swapping algorithm. Balances the trade-off between relevance and randomness.

```

Before Re Ranking:
Male Count: 11
Female_count: 9
After Re Ranking:
Male_Count: 10
Female_Count: 10

```

Figure 5: Fairness-greedy algorithm. Prioritizes fairness in the ranking to align within first few pages gender distributions.

Post-Algorithm Results: The fairness-greedy algorithm adjusted the distribution to a more balanced 10 male and 10 female images. **Analysis:** By prioritizing fairness in the initial pages of search results, the fairness-greedy algorithm significantly improves gender representation. This approach ensures that the first few pages, which are most visible to users, have a balanced distribution, directly addressing the visibility aspect of bias.

Comparative Analysis. A comparative analysis of the three algorithms reveals distinct strategies and outcomes in bias mitigation. The epsilon-greedy algorithm is effective in disrupting gender clustering, the relevance-aware swapping algorithm carefully balances relevance with bias mitigation, and the fairness-greedy algorithm prioritizes an equitable distribution, especially in the most visible search results. The findings from the application of these re-ranking algorithms underscore the potential to improve fairness in search engine results. The varying degrees of success across different algorithms highlight the importance of customizing bias mitigation strategies to the specific context and goals of the search engine. Further research is needed to refine these algorithms, perhaps by combining aspects of each to achieve both relevance and fairness. Additionally, longitudinal studies could assess the algorithms' impact on user experience and satisfaction.

5 LIMITATIONS:

One of the primary limitations of our study, which aligns with the concerns [5], is the binary perception of gender. The algorithms and the modified ResNet-10 architecture utilized in this study inherently categorize gender into male and female. This binary classification overlooks the spectrum of gender identities that extend beyond. Such a simplification not only misrepresents individuals but also

perpetuates the binary framework within which most AI systems currently operate. Moreover, the reliance on public census data as ground truth in the Fairness-Greedy algorithm introduces another potential source of bias. These datasets may not always reflect the evolving understanding of gender roles, which can lead to the reinforcement of outdated stereotypes.

6 FUTURE WORK:

In the pursuit of refining AI and its applications, there's significant work ahead. One priority is developing AI that recognizes a wider spectrum of gender identities, rather than just male or female. This not only means creating smarter, more sensitive algorithms but also asking people how they identify themselves to ensure the AI respects everyone's self-identified gender. Another key area is making AI's decisions clearer to everyone. If users can see how and why an AI labels images the way it does, they can also help it learn from its mistakes, like pointing out when it mislabels someone's gender. AI also needs to keep up with the times. Instead of using outdated statistics, AI should use the latest information to understand current gender roles better. There's also a need to consider the wider effects of AI on society. Are these systems fair to everyone? Do they accidentally reinforce stereotypes? Answering these questions might involve experts from many fields, like social scientists, alongside technologists. Lastly, it's not just about gender. AI should also strive to be fair regardless of someone's race, age, or any other aspect of their identity. And for all this to happen, we might need new laws or rules to guide the fair use of AI, ensuring it works well for everyone.

7 CONCLUSION:

Our research emphasizes the need for consistent, systematic strategies to combat gender bias in AI, moving beyond temporary fixes to more sustainable solutions. By implementing re-ranking algorithms like the Epsilon-Greedy, Relevance-Aware Swapping, and Fairness-Greedy algorithms, we've developed methods that not only improve gender representation in search engine results but also balance accuracy with diversity. These methods have been crucial in helping AI systems transition from merely reflecting existing biases to actively promoting a diverse representation of genders. For instance, our Fairness-Greedy Algorithm adjusts rankings based on real-world gender distributions in various professions, ensuring top images reflect a fair representation. This work is a step towards AI systems that are both fairer and more reflective of our diverse society, ensuring digital realms respect and acknowledge everyone's identity.

REFERENCES

- [1] *Baidu Search Engine*, 2024. Accessed: 2024-02-24.
- [2] *Naver Search Engine*, 2024. Accessed: 2024-02-25.
- [3] *Yandex Search Engine*, 2024. Accessed: 2024-02-22.
- [4] Facebook. *React Documentation for Web Development*, 2024. Accessed: 2024-03-12.
- [5] Yunhe Feng and Chirag Shah. Has ceo gender bias really been fixed? adversarial attacking and improving gender fairness in image search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 11882–11890, 2022.
- [6] Ruoyuan Gao and Chirag Shah. Toward creating a fairer ranking in search engine results. *Information Processing & Management*, 57(1):102138, 2020.
- [7] googlesearch-python 1.2.3 Google search apis. A python library for scraping the google search engine. In <https://pypi.org/project/googlesearch-python/>, 2022.
- [8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. volume abs/1512.03385, 2015.

- [9] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. 2014.
- [10] Matthew Kay, Cynthia Matuszek, and Sean A Munson. Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd annual acm conference on human factors in computing systems*, pages 3819–3828, 2015.
- [11] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: single shot multibox detector. volume abs/1512.02325, 2015.
- [12] Danaë Metaxa, Michelle A Gan, Su Goh, Jeff Hancock, and James A Landay. An image of society: Gender and racial representation and impact in image search results for occupations. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–23, 2021.
- [13] Jahna Otterbacher, Jo Bates, and Paul Clough. Competent men and warm women: Gender stereotypes and backlash in image search results. In *Proceedings of the 2017 chi conference on human factors in computing systems*, pages 6620–6631, 2017.
- [14] Pallets Projects. *Flask Documentation for Backend Development of the solution*, 2024. Accessed: 2024-03-01.
- [15] Dushyant Sharma, Rishabh Shukla, Anil Kumar Giri, and Sumit Kumar. A brief review on search engine optimization. In *2019 9th international conference on cloud computing, data science & engineering (confluence)*, pages 687–692. IEEE, 2019.
- [16] Ellen M Voorhees et al. The trec-8 question answering track report. In *Trec*, volume 99, pages 77–82, 1999.
- [17] Tianlu Wang, Jieyu Zhao, Mark Yatskar, Kai-Wei Chang, and Vicente Ordonez. Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5310–5319, 2019.
- [18] Fons Wijnhoven and Jeanna Van Haren. Search engine gender bias. *Frontiers in big Data*, 4:29, 2021.
- [19] wuanagana. res10_300x300_ssd_iter_140000_fp16.caffemodel. GitHub, 2020.
- [20] Tian Xu, Jennifer White, Sinan Kalkan, and Hatice Gunes. Investigating bias and fairness in facial expression recognition. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part VI* 16, pages 506–523. Springer, 2020.