Automated Perceived Gender Bias Pipeline in YouTube

Gizem Gezici*†
*Huawei Turkey R&D Center, Istanbul, Turkey
†Sabanci University, Istanbul, Turkey
gizem.gezici@huawei.com, gizemgezici@sabanciuniv.edu

Abstract—Students are increasingly using online materials to learn new subjects or to supplement their learning process in educational institutions. Issues regarding gender bias have been raised in the context of formal education and some measures have been proposed to mitigate them. However, online educational materials in terms of possible gender bias and stereotypes which may appear in different forms are yet to be investigated in the context of search bias in a widely-used search platform. As a first step towards measuring possible gender bias in online platforms, we have investigated YouTube educational videos in terms of the perceived gender of their narrators. We adopted bias measures for ranked search results to evaluate educational videos returned by YouTube in response to queries related to STEM (Science, Technology, Engineering, and Mathematics) and NON-STEM fields of education. For this, we propose automated pipeline to annotate narrators’ perceived gender in YouTube videos for analysing perceived gender bias in online education.

I. INTRODUCTION

Similar to the methodology in [1], in the context of this paper same educational queries are used for the bias analysis and the abbreviation of YVRP is used for the YouTube video result page. The gender of the narrator is similarly explored in this work, and the main aim is to label the videos according to the narrator’s perceived gender. Unlike [1], in this paper a given video is not labelled with a single gender label rather, with a probability distribution of male and female labels that could depict a more realistic perceived gender spectrum of a given video. Note that the perceived gender annotation will be automatically determined using the voice information of narrator(s) instead of manual annotation. For the detailed annotation procedure, please refer to our previous work [1].

Gender stereotypes are a common source of bias that emerge when an individual or a group is systematically treated favourably or unfavourably, referring to individual or group fairness respectively and there is a need to investigate gender representation in educational resources. In fact, the European Institute for Gender Equality states that gender stereotypes still exist in educational materials [2]. There are some guidelines on how to evaluate diversity in educational materials, for example Michigan in the United States issued a report in 2020 as a guidance for the experts in evaluating instructional materials in terms of bias [3].

YouTube states that they audit their machine learning systems to avoid cases leading to gender discrimination [4]. However, this does not guarantee that the returned videos are not biased towards a specific gender. In this study, our aim is to investigate educational videos returned by YouTube in terms of possible gender bias via objective measures. We focus on group fairness since we investigate if the online materials are affected by societal stereotypes about gender in the context of education. Moreover in group fairness, we specifically focus on statistical parity, demographic parity or more generally known as equality of outcome, i.e. given a population divided into groups, the groups in the output of the system should be equally represented. In the scope of this work, equality of outcome is a more appropriate standard since we require equal gender representations in results. Our main aim in this study is to detect bias with respect to equality of outcome using the perceived gender of narrators in videos returned by YouTube in response to educational queries comprising of keywords regarding some educational field. For this purpose, we use educational queries that are derived from the course modules of STEM and NON-STEM fields.

Our contributions can be summarised as follows:
1) We propose an automated pipeline to annotate narrators’ genders in YouTube videos which could be used for large-scale bias analysis.
2) We present two fine-grained measures of bias which are explained in Section II-B in detail that represent the gender spectrum of a given video with a probability distribution which is a more realistic representation, and generate bias values which are symmetric and easy to interpret.
3) We implement our fine-grained bias measures to investigate possible perceived gender bias as well as compare the relative bias for educational searches in YouTube about different majors from STEM and NON-STEM fields.
4) We further investigate the source of perceived gender bias in YouTube for educational searches owing to the proposed automated pipeline.

II. GENDER BIAS EVALUATION METHODOLOGY

This section describes the methodology for evaluating perceived gender bias without a binary gender assumption. The bias measures proposed in [1] are adapted to using male/female probability values instead of single perceived gender labels. Further, a similar evaluation procedure presented in [1] is fulfilled for identifying potential bias as well as
tracking the source of bias (if applicable) associated with those adapted measures.

A. Preliminaries and Research Questions

Using an automated approach to obtain the perceived gender probability distributions of videos through voice information, this work mainly aims to following research questions. The first research question is:

RQ1: On a perceived male-female gender spectrum, does YouTube return biased YVRPs in response to various educational queries using more fine-grained measures?

The second research question is:

RQ2: Is there a significant difference in perceived gender bias that is computed in a more fine-grained manner in YVRPs returned in response to STEM vs. NON-STEM educational queries?

The third research question is:

RQ3: Do different cut-off values affect the existence of perceived bias and magnitude of bias difference by using more fine-grained measures between STEM and NON-STEM fields?

The fourth research question is:

RQ4: Do different cut-off values affect the magnitude of perceived bias of STEM and NON-STEM fields separately that is measured in a more fine-grained manner?

The last research question which is totally a new question that could not be answered in our previous work [1]. In the context of this paper, the source of bias (if exists) is investigated. If the bias measured in the full YVRPs is consistent with the top video search results, i.e. especially in top-3, top-10 since these search results attract users’ attention the most, then it can be inferred that the bias comes from the data itself. If there are some differences between those bias results, then it can be inferred that the bias comes from the data itself. The ranking algorithm could also be blamed. Note that the data and ranking algorithm both could be responsible for the biased YVRPs.

The fifth research question is:

RQ5: What is the source of bias (if exists), does it come from the input data, or the ranking algorithm?

B. Measures of Bias

Let \( q \in Q \) be the set of educational queries about the majors in STEM and NON-STEM fields. When a query \( q \in Q \) is issued to YouTube, YouTube returns a YVRP \( r \). The probability value associated with the perceived gender of the \( i \)-th retrieved video \( r_i \) with respect to \( q \) is defined as \( \text{prob}_{G_m}(r_i) \) for male and \( \text{prob}_{G_f}(r_i) \) for female. For reference, Table I shows a summary of all the symbols, functions and labels used throughout this paper.

In the scope of this paper, similar to Section I, the main aim is to satisfy group fairness criteria of equality of outcome where male and female genders should be equally represented in YVRPs. Thus the perceived gender bias is measured as the difference between the representation of male and female genders.

Formally, the perceived gender bias in a YVRP \( r \) is measured as follows:

\[
\Delta f(r) = f_{G_m}(r) - f_{G_f}(r) \tag{1}
\]

For the function \( f(r) \), two bias measures that are proposed in [1] are adapted to using probability scores instead of single perceived gender labels in the scope of this paper. Note that the videos annotated with \( G_{\text{not rel}} \) and \( G_{N/A} \) are initially discarded before the bias score computations. The two adapted measures of representation and exposure are denoted by \( \text{Rep}_{\text{prob}}@n \) and \( \text{Exp}_{\text{prob}}@n \) respectively. The first adapted measure of bias, \( \text{Rep}_{\text{prob}}@n \) which computes a bias score using probability values associated with the perceived gender label of male as follows:

\[
\text{Rep}_{\text{prob}}@n = \frac{1}{\text{prob}_{mf}} \sum_{i=1}^{n} \text{prob}_{G_m}(r_i) \tag{2}
\]

Note that \( \text{Rep}_{\text{prob}}G_f@n \) is computed in the same way. The following equation by substituting Eq. (2) in Eq. (1):

\[
\Delta_{\text{Rep}_{\text{prob}}@n}(r) = \frac{1}{\text{prob}_{mf}} \sum_{i=1}^{n} (\text{prob}_{G_m}(r_i) - \text{prob}_{G_f}(r_i)) \tag{3}
\]

Since the first bias measure of \( \text{Rep}_{\text{prob}}@n \) has a weak sense of rank information, i.e. all the positions contribute to bias score in an equal manner, the second measure is presented by adapting \( \text{Exp}_{\text{prob}}@n \) that was proposed in [1] to using probability scores. The second adapted measure of bias, \( \text{Exp}_{\text{prob}}@n \) which computes a bias score in terms of exposure, using probability values associated with the perceived gender label of male as follows:

\[
\text{Exp}_{\text{prob}}G_m@n = \sum_{i=1}^{n} \frac{1}{\log(i + 1)} \left( \frac{\text{prob}_{G_m}(r_i)}{\text{prob}_{G_m}(r_i) + \text{prob}_{G_f}(r_i)} \right) \tag{4}
\]

Note that \( \text{Exp}_{\text{prob}}G_f@n \) is computed in the same way. The following equation is obtained by substituting Eq. (4) in Eq. (1):

\[
\Delta_{\text{Exp}_{\text{prob}}@n}(r) = \sum_{i=1}^{n} \frac{1}{\log(i + 1)} \left( \frac{\text{prob}_{G_m}(r_i) - \text{prob}_{G_f}(r_i)}{\text{prob}_{G_m}(r_i) + \text{prob}_{G_f}(r_i)} \right) \tag{5}
\]

The scores of the proposed measures are easy to interpret, for a given ranked list, the scores of two gender groups sum up to 1. If the bias scores are interpreted with respect to the equal representation using \( \text{Rep}_{\text{prob}}@n \), then it can be inferred which gender group is more/less represented than the desired representation. Same holds true for the exposure measure, \( \text{Exp}_{\text{prob}}@n \) which determines if a gender group is more or less exposed than the desirable situation of the equal
exposure. For interpreting the results, if the value of 0.5 which is the desired case, is subtracted from the measure scores of male and female gender for a given list, then the remaining bias scores of each gender group are symmetric. Same holds for the exposure measure. Additionally, these adapted bias measures are expected to compute smoother and more realistic bias scores owing to the probability scores of the perceived gender groups instead of single labels which are too deterministic for annotating real datasets.

After the computation of adapted representation and exposure bias scores, the mean bias (MB) and mean absolute bias (MAB) of these measures can be further computed over a set of queries in the dataset to aggregate the bias results. MB score of STEM field computes a mean value over all the STEM queries’ scores for the corresponding measure, whereas MAB computes a mean value over all the absolute value of the measure scores for the STEM queries. Note that MB shows towards which perceived gender the results are biased and MAB solves the limitation of MB if different queries have bias contributions with opposite signs and cancel each other out. Thus, MB and MAB measures are complementary for aggregating the results and interpreting those results in a proper way.

Please note that in the context of this paper, the probability scores correspond to each perceived gender label, i.e. \( \text{prob}_{G_m}(r_i) \) and \( \text{prob}_{G_f}(r_i) \) for male and female respectively, of a given video is computed merely based on the narrators’ perceived gender by using voice information of the narrators. For the details about the automated annotation procedure, please refer to Section III-B. Since the probability scores are leveraged, there is no gender binary assumption throughout this paper.

C. Quantifying Bias

Using the measures of bias defined in Section II-B, the perceived gender bias of the STEM and NON-STEM fields is first measured in YVRPs returned in response to the educational queries in various majors, and then a comparative evaluation is fulfilled.

- **Collecting YVRPs.** The same educational query list was used to crawl YVRPs, for the complete query list, please refer to \( \square \). For each query, top-200 video results returned by YouTube UK in incognito mode was crawled by using the YouTube Data API v3 \( \square \). Note that the data collection process was done in a controlled environment such that the queries were sent to YouTube via the YouTube API by avoiding long time-lags. After the crawling, all the video results related to the majors in both STEM and NON-STEM fields, they were automatically assigned with probability scores of each perceived gender group using the voice information of the narrators.

- **Bias Evaluation.** The bias scores are computed for every YVRP with two adapted bias measures with four different cut-off values: \( \text{Rep}@n \) and \( \text{Exp}@n \) for \( n = 3, 10, 20, \) and 200, e.g. full list where \( |r| = 200 \). Then, the results are aggregated using the MB and MAB. Additionally, first the existence of bias for each field is examined, then the bias results of STEM and NON-STEM fields with different measures and cut-off values are compared. Subsequently, the impact of different cut-off values is investigated on bias scores of STEM and NON-STEM fields. Finally, the source of bias is investigated by comparing the top video search results with the full list, i.e. top-3, top-10 vs. top-200.

- **Statistical Analysis.** To identify whether the bias measured is not due to noise, a one-sample t-test is applied.
Note that since the sample size is sufficiently large (>30), according to the central limit theorem the sampling distribution is considered normal. If this hypothesis is rejected, hence there is a significant difference and it can be claimed that the YVRPs of the evaluated field, STEM or NON-STEM is biased. The difference in bias measured across the two fields is further compared using a two-tailed independent \( t \)-test. In addition to the statistical significance, namely p-values, effect sizes are also reported using Cohen’s \( d \). Statistical significance helps to examine whether the findings show systematic bias or they are the result of noise, whereas effect sizes provide information about the magnitude of the differences. Thus, both p-values and effect sizes provide complementary information for the interpretation of the results.

Apart from these, to investigate the effect of different cut-off values on bias results in the same field, STEM or NON-STEM, a two-tailed paired \( t \)-test is computed since the same query set is examined for different cut-off values and for two different annotation models. Moreover, Bonferroni correction is further applied for multiple hypothesis testing since there are 60 hypotheses in total in the context of cut-off value and annotation model analyses. Thus, without the Bonferroni correction, with the significance level, \( \alpha = .05 \) and 60 hypotheses, the probability of identifying at least one significant result due to chance is around 0.95 which means that the results could be misleading. Hence, the Bonferroni correction is also applied for more reliable results. Note that for the significance level where \( \alpha = .05 \), and with the Bonferroni correction new \( \alpha = .0008 \). Thus, Bonferroni correction rejects the null hypothesis for each p-value \( (p_i) \) if \( p_i \leq .0008 \) instead of .05. For the significance level where \( \alpha = .01 \), with the Bonferroni correction new \( \alpha = .0002 \).

### III. Experimental Setup

In this section, first dataset information, then the annotation procedure, and lastly the perceived gender bias results will be provided based on the proposed method as described in Section II. In addition to the existence of bias, source of perceived gender bias is investigated as well.

#### A. Dataset

The same educational query list was used to crawl the dataset. For the crawling, YouTube Data API v3 was utilised. Note that crawling with a Python implementation using the YouTube Data API v3 was fast enough that it does not create noticeable time lags between queries which could affect the bias analysis. Initially, several API keys were created for free. Then, using these API keys the location, i.e. \( \text{regionCode} \) in the API document, was set to the UK and YVRPs were crawled automatically. Yet, YouTube Data API v3 has some limitations as follows. First, one can crawl a limited number of YVRPs for each generated API key; the quota is based on the information crawled for each API request. Thus, it is important to crawl only sufficient information while using the YouTube Data API.

For the detailed quota information, one can use the official page. Second, the YouTube Data API v3 returns 50 YVRPs in total for each API request; this is the maximum number that could be retrieved using the official API per request. Therefore, one needs to find a workaround to crawl more than 50 YVRPs per query. For this purpose, the YouTube Data API v3 provides a field of \( \text{nextPageToken} \) that denotes a unique ID for the next page of the YVRPs of the current query. Then, this \( \text{nextPageToken} \) could be assigned to the \( \text{pageToken} \) field of the YouTube Data API v3 request while crawling the next page of the video search results of the same query.

Based on the aforementioned information, using the YouTube Data API v3 with the same educational query list in \( \text{YVRP} \), in total 200 YVRPs were crawled for each query. Similarly, in the scope of this work, since personalised search might complicate the bias analysis, the analysis was designed in unpersonalised search settings, i.e. there was no user information included in the API requests. The main reason of using automated crawling and annotation procedures in this paper is to investigate the source of bias in search results as well. Hence, 200 YVRPs were obtained with the assumption that these search results could be the representative of the full video search results for the corresponding query that could be used to track the source of bias. Since crawling more YVRPs require huge processing time, especially in the annotation phase which uses deep learning-based automated models, 200 YVRPs were selected for the source of bias analysis. To crawl the 200 YVRPs for each query, four API requests were sent to the YouTube Data API v3 using the \( \text{nextPageToken} \) information. In this way, 200 results were obtained for all the educational queries, except the query of \( \text{Capital labour and power: Britain 1707-1939} \) in the NON-STEM major of \( \text{Politics} \). Only for this query, YouTube Data API v3 returned 67 YVRPs in total and changing punctuations in the query etc. did not change the retrieved search results by YouTube. Please note that the educational query list was not modified on purpose in order to obtain comparative evaluation results with our previous work and not to inject personal bias.

#### B. Annotation Procedure

The automated annotation procedure has two essential steps. First a video in a YVRP is annotated based on its relevancy with respect to the given query as relevant, not-relevant. If the video is relevant to the given query, then the video is annotated with a probability distribution of \( \text{male} \) and \( \text{female} \) genders based on the perceived gender of narrator(s) through voice information instead of labelling each video with a single perceived gender label as fulfilled in \( \text{YVRP} \). Note that before these two main annotation steps, videos with the following properties were discarded.

- **Short Videos**: The videos that are shorter than 20 seconds.
- **Unavailable Videos**: The videos that has been removed by the user or from YouTube because of copyrights.
- **Restricted Videos**: The videos that require a user login to watch either because it is a private video or because of age restriction.

[https://developers.google.com/youtube/v3/determine_quota_cost](https://developers.google.com/youtube/v3/determine_quota_cost)
After discarding these videos, the first step of the annotation procedure was applied.

1) Relevance Annotation: After the crawling, to automatically detect the relevance label of a given video with respect to the corresponding query, a document similarity approach was implemented. YVRPs were converted into textual contents which were then used to measure document similarity. The main idea here is to utilise the dataset in [1] that had already been annotated with the relevance labels. The approach of measuring document similarity combines the following three models to encode documents: Term-frequency multiplied by inverse document-frequency (tf*idf) [3], Universal Sentence Encoders (USE) [4] and Sentence-BERT (SBERT) [8] using SentenceTransformers [5] Python framework. In addition to these, a jaccard similarity measure [9], which is mainly based on the number of common words between two documents, was also implemented. Yet, the jaccard measure did not work well so it was discarded from the analysis.

After the implementation of these models, threshold values of each model were determined experimentally. For evaluating the capability of these models and tuning the threshold values on the textual contents of YVRPs, the dataset in [1] was used as mentioned above. Then the results of the aforementioned three models of tf*idf, USE, and SBERT were merged for measuring document similarity. To automatically label the relevancy of the YVRPs that were crawled with the data crawling procedure as described in Section III-A, these steps were fulfilled:

(i) Using the videoID, title, description, and subtitles of each video in the YVRPs of were crawled.

(ii) Then, title, description and subtitles of each video were concatenated; each video was represented with this concatenated textual content.

(iii) A preprocessing phase of removing numbers, punctuation, making lowercase, and lemmatisation was applied.

(iv) The three models of tf*idf, USE, and SBERT were applied on the preprocessed textual contents of the YVRPs that were previously crawled and annotated in [1].

(v) Given the true relevancy label of a given document (video), experimentation was fulfilled with model representations of the given document and different threshold values for each model.

(vi) The threshold values of 0.1, 0.5 and 0.5 were determined for tf*idf, USE, and SERT respectively. If the computed document similarity score is below a threshold, then the document is labelled as not-relevant for the corresponding model.

To automatically annotate a given video in the rest of the videos with a relevance label, three different models were deliberately used. After computing document similarity scores, these scores were reviewed as well. For the final decision about the relevancy of a given video, the document similarity scores of the three models were expected to be consistent. This means that a more conservative approach was taken, i.e. a given document/video is not-relevant if these three models agree, in order not to lose the relevant documents.

Based on the reviewing of the document similarity scores obtained from these three models, it was observed that the combined approach worked sufficiently well. The combined approach has the capability of detecting non-English as well as out-of-context videos. For instance, YouTube returned some videos related to Python context managers for the query/course module of Management in context in Public Relations major of NON-STEM field and the combined approach detected labelled those videos as not-relevant. Additionally, the combined model detected some non-English videos for the query/course module of Urban sociology in Sociology major of NON-STEM field. Note that in the scope of this paper, an additional relevance label of N/A was not used for non-English videos and videos without a narrator. Instead, non-English videos were labelled together with out-of-context videos as not-relevant. The videos with no narrator were handled during the perceived gender annotation in Section III-B2.

2) Perceived Gender Annotation: As the second phase of the annotation procedure, if a given video is relevant then it was annotated automatically with perceived gender information by computing its probability distribution of male
and female gender labels based on the audio of the given video. This phase is composed of two main steps. First, the given audio was classified into the segments of speech, music, or noise using the model of inaSpeechSegmenter [10]. Note that the audio segment of music or noise correspond to no narrator case, i.e. this refers to the label of N/A. Subsequently, the inaSpeechSegmenter and Feed-Forward Gender Detector [11] models were used independently to detect perceived gender on the speech segments. Finally, male (female) ratio was computed by measuring the time of the audio segment that is annotated with male (female) perceived gender label divided by the time of the full speech audio segment. The perceived gender annotation procedure is displayed in Figure 1.

inaSpeechSegmenter proposes a gender detection processing pipeline which is composed of three main parts. The first sub-module of a Speech/Music segmenter based on Convolutional Neural Networks (CNN) [12] is responsible for discarding music and empty segments. Subsequently, features corresponding to speech segments are extracted using a common extraction framework. A simple energy threshold is utilised to discard frames with low energy. Lastly, Gaussian Mixture Models (GMMs), i-vectors, i.e. compact vector representation of speech utterance, and CNN systems are then leveraged to classify the remaining speech segments into male and female excerpts. The inaSpeechSegmenter has been trained on the INAs Speaker Dictionary [11] which contains about 32000 excerpts of 1780 male (94 hours) and 494 female (27 hours) speakers. This audiovisual corpus was annotated with a semi-automatic labelling protocol. For more details about inaSpeechSegmenter, please refer to the original paper [10].

The second model of the feed-forward gender detector is a deep feed-forward neural network of five hidden layers, i.e. 0.3 dropout rate after each dense layer, was presented in this tutorial [1]. The dataset that was used in the second model is Mozilla’s Common Voice Dataset [7] which is a corpus of speech data read by users on the Common Voice Website [8]. Before using the dataset, the dataset was balanced as the number of male samples equal to female samples, i.e. in total 67K samples with equal number of male/female samples, to prevent the model to favour one particular gender. In addition, for feature extraction this second model utilises Mel Spectrogram [9] extraction technique to obtain a compact vector representation of length 128. For more details, please refer to the tutorial and the implementation of the model.

For evaluating the capability of the aforementioned two automated models, the dataset of VoxCeleb [12] was used. VoxCeleb contains audio and video of short clip extracted from the interviews on YouTube. The dataset includes speakers from a wide range of ethnicities as well as ages. In terms of gender, the dataset is composed of 3682 male speakers (61%), and 2311 female (39%). Note that during the evaluation phase, only the audio clips were used since in the scope of this work, YouTube videos are annotated with probability distributions using only the audio information in a given video. Since the evaluation dataset is also from YouTube, model evaluation results for binary classification could be a good indicator in terms of the models’ annotation capability. Based on the evaluation results shown in Table 1, it is observed that inaSpeechSegmenter outperforms Feed-Forward Gender Detector by a large margin – inaSpeechSegmenter gives around 20% higher F1-scores for the female class, while around 10% for the male class.

Note that the length of the educational videos crawled from YouTube might vary from couple of minutes to multiple hours. Hence, applying two deep learning-based models to get annotations might take a huge amount of time. To speed up the annotation procedure, rather than annotating the full video, a group of samples were selected from each video. Moreover, most of these educational videos start with a musical intro and end with a musical outro. Thus, in order to make the sampling more focused on the video content, the sampling was initiated 10 seconds after the beginning and terminated 10 seconds before the ending. Then, for the sampling of each video, i.e. after discarding both the first and last 10 seconds, first 10 seconds of each video minute was taken as a sample. Finally, on each sample the annotation steps as denoted in Figure 1 were executed. In this way, male/female probability distributions of each video were computed using the corresponding sample-level annotations of the aforementioned two models. Instead of perceived single labels, probability distributions are expected to provide smoother and more realistic bias results in comparison to the results in [1].

| TABLE II: Evaluation Results on VoxCeleb |
|-----------------------------------------|
| **Model**                          | **Gender** | **Precision** | **Recall** | **F1-score** |
|-------------------------------------|------------|---------------|------------|--------------|
| Feed-Forward Gender Detector       | Female     | 0.79          | 0.79       | 0.79         |
|                                    | Male       | 0.87          | 0.87       | 0.87         |
| inaSpeechSegmenter                  | Female     | 0.92          | 0.94       | 0.93         |
|                                    | Male       | 0.96          | 0.95       | 0.96         |

REFERENCES

[1] G. Gezici and Y. Saygin, “Measuring gender bias in educational videos: A case study on youtube,” arXiv preprint arXiv:2206.09987, 2022.
[2] EU. (2017, 1) Education. [Online]. Available: https://eige.europa.eu/gender-mainstreaming/policy-areas/education
[3] C. Center, “Tools and Guidance for Evaluating Bias in Instructional Materials,” 2020.
[4] YouTube. (2018) Youtube’s impartiality practices - how does youtube work? [Online]. Available: https://www.youtube.com/howyoutubeworks/our-commitments/preventing-bias
[5] https://github.com/x4nth055/gender-recognition-by-voice
[6] https://en.wikipedia.org/wiki/Convolutional_neural_network
[7] https://www.robots.ox.ac.uk/~vgg/data/voxceleb/
[8] https://github.com/x4nth055/gender-recognition-by-voice
[9] https://www.jetbrains.com/article/gender-recognition-by-voice-using-tensorflow-in-python
[10] https://www.kaggle.com/datasets/mozillaorg/common-voice
[11] https://en.wikipedia.org/wiki/Mel
[12] https://www.robots.ox.ac.uk/~vgg/data/voxceleb/
[5] S. G. Kwak and J. H. Kim, “Central limit theorem: the cornerstone of modern statistics,” Korean Journal of Anesthesiology, vol. 70, no. 2, p. 144, 4 2017. [Online]. Available: pmc/articles/PMC5370305/?report=abstract

[6] P. Sedgwick, “Multiple significance tests: the bonferroni correction,” BMJ, vol. 344, 2012.

[7] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, and C. Tar, “Universal sentence encoder,” arXiv preprint arXiv:1803.11175, 2018.

[8] N. Reimers and I. Gurevych, “Sentence-bert: Sentence embeddings using siamese bert-networks,” arXiv preprint arXiv:1908.10084, 2019.

[9] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, “Using of jaccard coefficient for keywords similarity,” in Proceedings of the international multiconference of engineers and computer scientists, vol. 1, no. 6, 2013, pp. 380–384.

[10] F. Salmon and F. Vallet, “An effortless way to create large-scale datasets for famous speakers.” in LREC, 2014, pp. 348–352.

[11] D. Doukhan, J. Carrive, F. Vallet, A. Larcher, and S. Meignier, “An open-source speaker gender detection framework for monitoring gender equality,” in Acoustics Speech and Signal Processing (ICASSP), 2018 IEEE International Conference on. IEEE, 2018.