

Deepfake proposal 1.1

by Muhammad Talha Hassan Mehmood

Submission date: 03-Oct-2023 09:57PM (UTC-0700)

Submission ID: 2185170645

File name: deepfake_proposal_for_plag.docx (392.45K)

Word count: 1774

Character count: 9992

Abstract

This project aims to develop an application that can generate voice clones based on provided text using a trained model. Many Text-to-Speech models exist, they are limited to specific voices [1]. Our focus is on training a model with an English dataset to replicate a particular person's voice. Once trained, the model can produce Deepfake voices for the provided text, offering advantages in education, cost-effective advertising, and the media industry. However, it's essential to acknowledge the potential for misuse, especially in serious crimes and impersonation, necessitating subsequent efforts for detection and prevention. This project's primary objective is Deepfake text to speech creation, with ethical and security considerations in mind. [2]

1 Introduction

This project is an **industrial project** and industry name is **Expert System Solution**. Deepfake is a combination of words Deep which is Deep learning and Fake. Deep Learning is basically Artificial Intelligence based Learning Architecture which is trained on Dataset and give you a desire results. Creating something fake using Deep Learning is basically a Deepfake. While the act of creating fake content is not new, deep fakes leverage powerful techniques from machine learning and artificial intelligence to manipulate or generate visual and audio content that can more easily deceive. Deep fakes started impacting society in various ways like manipulation of facial appearance, Voice Cloning of famous Personalities, Fake Text Generation. [3]

Deep learning has made significant progress in recognizing actions in videos and Voice in real-time, and it no longer requires expensive and powerful servers as it used to. Thanks to improvements in hardware and software, we can now use these deep learning models on everyday devices like mobile phone. This development has opened exciting possibilities for creation of Deepfake Content, which our project aims to leverage fully. The most common deep learning-based machine learning techniques for producing Deepfakes include training generative neural networks designs like auto encoders or generative adversarial networks (GANs). [4]

Deepfakes are manufactured works of art in which the resemblance of a different person is used to replace a real person in a photograph, video and Voice. Now days, it is use in News Media as India channel Odisha TV launches the Deepfake audio and animation-based News Anchor. [5] Ukraine President Deepfake video goes viral which is in support of Russia cause a chaos in Public. [6] Now days, Deepfake is very much in use, mostly for Negativity in the society such as Harassment, False Information, the use of deep fakes since they began in 2017 has been shocking. A huge increase in articles related to Deepfake have been noticed from 2018 and in 2020, it was expected that around 730 articles will be publish related to Deepfake but in reality at end of year number was way more than that around 1323 [7], which shows the interest of people in exploring this field.

This Project aims the creation of Deepfake text to speech on a single person dataset, for this purpose considered datasets given below. In future, this project will provide us help in Deepfake detection and prevention as well as expanding the scope for multiple person's Deepfake and personal voice Deepfake.

2 Success Criteria

Our project's success is measured by achieving top-tier accuracy in creating Deepfake voices from text and ensuring efficient real-time performance with minimal power consumption, making our solution practical for various applications. Initially, our focus lies in generating Deepfake voices for specific individuals, using their datasets as a foundation. In the future, we expand our capabilities to create voices for a wider array of people, enhancing our project's versatility. Additionally, a user-friendly website is developed for seamless Deepfake voice generation and use. This platform facilitates responsible Deepfake technology utilization and supports voice cloning detection efforts, promoting ethical use.

3 Related work

3.1 Resemble AI

Resemble AI is an online platform which is used to create audio deep Fake provide text to speech generation over 60+ Languages. Also provide the deep Fake voice cloning detection services. Moreover, provide AI based audio editing services. This platform provide paid services. [11]

- i. Voice Cloning text to speech
- ii. Provide Voice Localization with over 60+ Languages
- iii. Neural Audio Editing
- iv. Real time audio deep fake detector

3.2 LOVO AI

LOVO is a hyper realistic AI voice generator which provides text to speech, Voice Cloning and can generate voice in more than 50 Languages. But it does not provide voice cloning on personal voice means Localization. It provides paid services. [12]

3.3 Audio deep fakes: A survey

It's a Research Paper published by researchgate.net by **Zahra Khanjani, Gabrielle Watson and Vandana P. Janeja** [7].

This research paper give a detail understanding related to Deep Fake audio as it covers some day to day life example of deep fake audio, Overview of deep learning and Artificial intelligence used in creation of Deepfake, provide how to build a Deepfake audio, provides list of pre trained Models and architectural working for the Deep fake text to speech Application:

- i. Real Time Deep Fake Audio Examples
- ii. Deep Learning and AI
- iii. Development and pre trained Models
- iv. Deep Fake Audio Architectural Explanation

- v. Explain the Concepts of Encoder and Decoder Architecture.
- vi. Advancement and Detection Techniques

3.4 Available Datasets:

- **LJ speech :**
LJ speech is a English speech dataset of a single speaker having short clips based on book passage reading consist of more than 13,000 audio clips. [8]
- **RyanSpeech:**
Ryan speech is a public dataset having 10 hours of audio material of a professional male voice actor's speech recorded at 44.1 kHz. [9]
- **LibriTTS:**
Libri TTS is very popular multi-speaker English dataset having 585 hours of speech audios at 24 kHz sampling rate. Need to customize it according to this project. [10]

4 Project Rationale

The rationale for this project lies in the increasing prevalence and potential misuse of Deepfake audio technology in today's digital age. As technology continues to evolve, the ability to manipulate and generate audio content with remarkable realism has raised critical ethical and security concerns. The misuse of Deepfake audio in legal proceedings and other domains threatens trust and integrity, necessitating a proactive approach to understand, address, and responsibly manage this technology. This project strives to develop solutions that safeguard against misuse and promote the ethical use of Deepfake voice cloning in an ever-advancing technological landscape.

4.1 Aims and Objectives

The aim of this project is to develop a Deepfake text-to-speech system for voice cloning, initially focusing on cloning the voice of one specific person and subsequently expanding to a broader range of individuals.

1. Achieving state-of-the-art accuracy in our model training on custom datasets.
2. Ensuring that the voice cloning can deceive the audience
3. State of the art custom dataset of a specific person which can to develop a realistic Deepfakes.

Objectives:

1. Create a Deepfake voice generation model capable of producing convincing audio from text input.
2. Address the ethical concerns associated with voice cloning, particularly in legal and law enforcement contexts.
3. Improve the accuracy and reliability of tools for detecting Deepfake audio.

4. Construct a comprehensive dataset for training Deepfake voice generation models.
5. Lay the groundwork for future research in Deepfake voice detection.
6. Promote responsible and ethical use of Deepfake voice technology.

4.2 Scope of the Project

The project's initial scope involves developing a Deepfake text-to-speech system for cloning the voice of a single individual, for which considering online dataset. However, the project aims to broaden its scope to include the capability to clone voices of a wider range of people in the future.

The project also encompasses efforts to create a user-friendly website that facilitates the generation and responsible use of Deepfake voices. Additionally, it contributes to the field of Deepfake voice detection, with the potential for further research and development in this area. Ultimately, the project's scope emphasizes responsible management of voice cloning technology and its implications for various domains, including law, security, and society as a whole.

5 Proposed Methodology and Architecture

5.1 Model Architecture

In the development of our cutting-edge Deepfake Text-to-Speech (TTS) system, we will harness the power of advanced machine learning techniques, specifically focused on cloning the voice of a single individual. The approach involves utilizing state-of-the-art models available through resources like Hugging Face, such as 'suno/bark' and 'speech tts,' as a foundation. First collecting a high-quality sample of the target person's voice, which will serve as the basis for training our customized TTS model.

Similar to the approach proposed by their MoViNet-based system, we recognize the value of transfer learning. By fine-tuning the pre-trained TTS models with the unique vocal characteristics of our chosen individual, the target to create a highly accurate and natural-sounding replica of their voice. This tailored TTS system will be capable of generating speech that mimics not only the linguistic style but also the distinct tonal nuances of the target person.

The objective is to achieve remarkable results in voice cloning, enabling applications in voice assistant technology, voiceovers, and personalized content creation. This approach aligns seamlessly with the trend towards edge computing, as our optimized TTS model will be designed for real-time performance on a variety of edge devices, ensuring a seamless and efficient user experience in various applications.

5.2 Frontend & Backend

The Deepfake voice cloning project is strategically planned for comprehensive development and subsequent deployment as a sophisticated web application. Within this well-thought-out framework, the trained model will be strategically deployed on the backend infrastructure. This deployment will be facilitated by the creation of a powerful API, with

flexibility in choosing between the Node.js framework Express or the Python framework Django, to cater to specific project needs.

Once this crucial backend integration is successfully established, the project will seamlessly transition to the frontend component's development phase. This frontend component is thoughtfully crafted using React JS, ensuring a seamless and user-friendly interface that harmonizes perfectly with the backend infrastructure, thus creating a unified and efficient Deepfake voice cloning application.

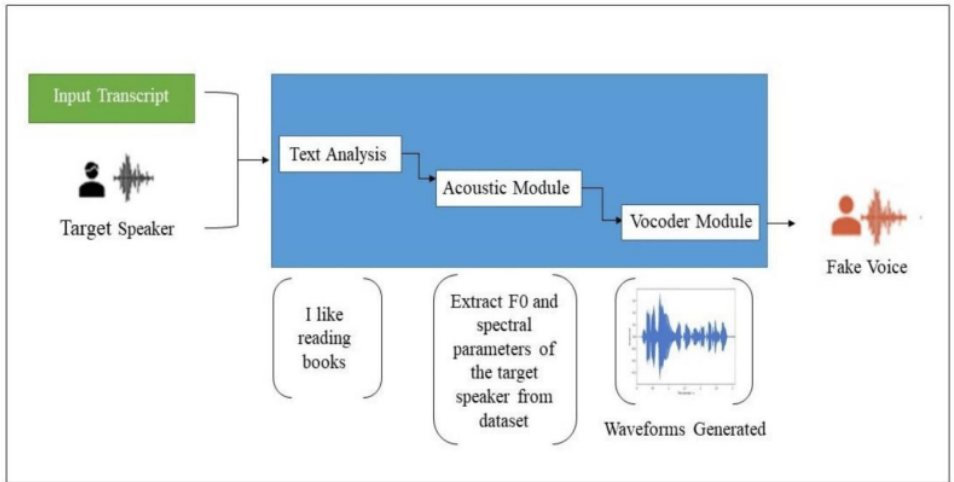


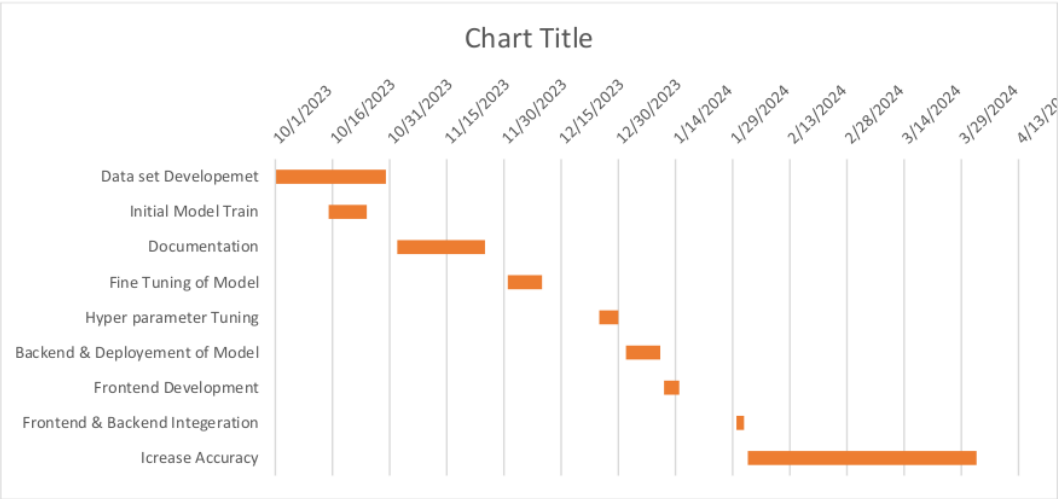
Figure 1 System Architecture

6 Individual Tasks

Team Member	Activity	Tentative Date
Hassan Mahmood	Dataset Development	October 15
Muhammad Talha	Refine dataset and tune according to selected model	October 30
Muhammad Talha, Hassan Mahmood	Documentation	November, 30
Muhammad Talha	Train initial demo of Model	December 15

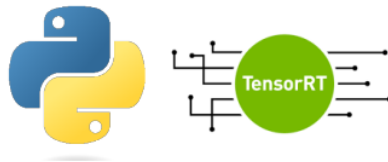
Hassan Mahmood	Fine Tuning of pre trained Model	December 15
Hassan Mahmood, Muhammad Talha	Hyper Parameter Tuning for better accuracy.	December 30
Muhammad Talha	Backend and deployment of Model	January 10
Hassan Mahmood	Frontend Development	January 30
Hassan Mahmood, Muhammad Talha	Frontend and Backend integration	January 30
Hassan Mahmood, Muhammad Talha	Increase Performance Metrics	March 30

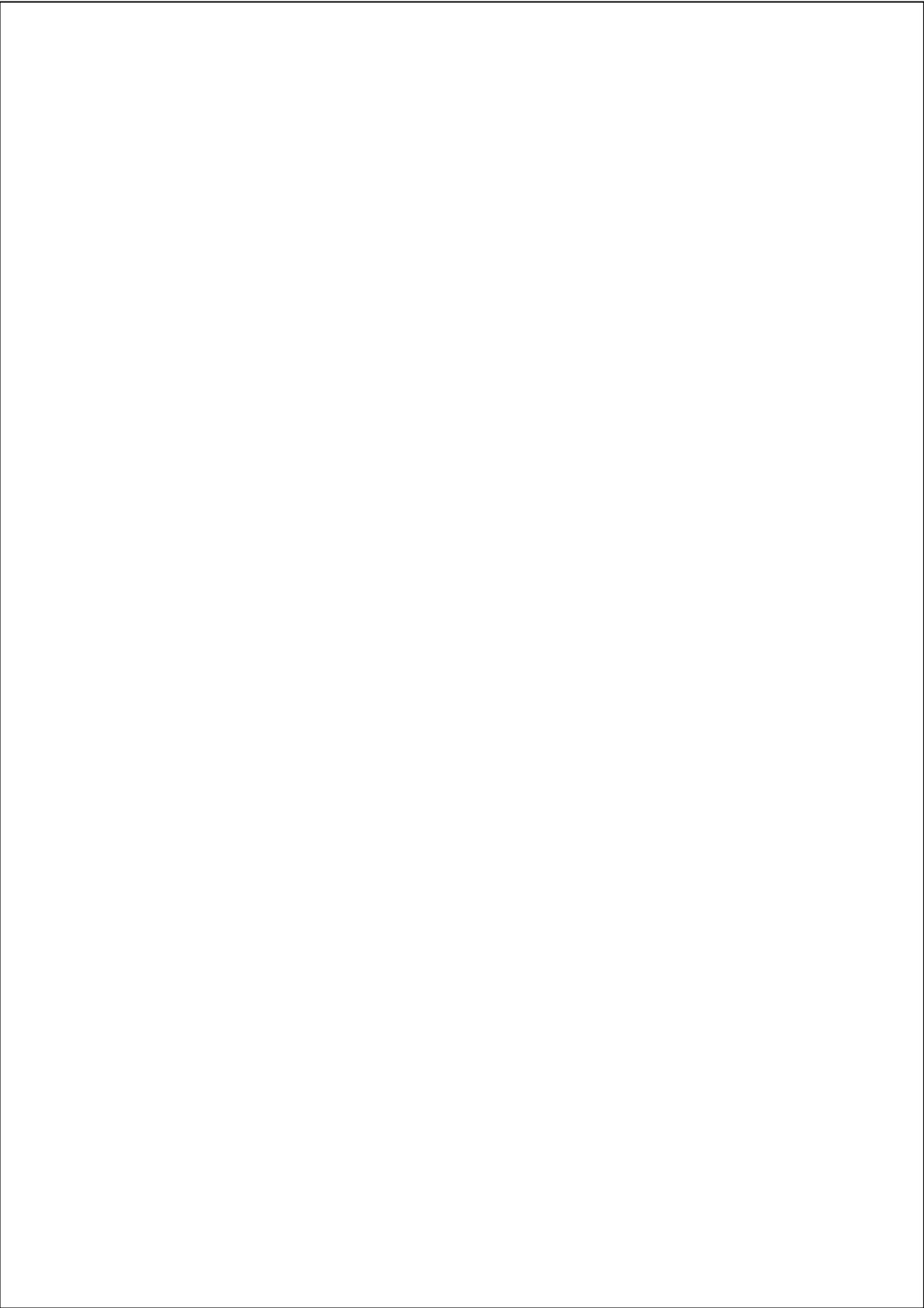
7 Gantt Chart



8 Tools and Technologies

- Google Colab
- Python
- TensorFlow / Pytorch
- Hugging Face
- NodeJS
- Express JS
- React JS
- Mongo DB





Deepfake proposal 1.1

ORIGINALITY REPORT

7%

SIMILARITY INDEX

4%

INTERNET SOURCES

0%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

1

frrosx.ballonskaufen.de

Internet Source

2%

2

Submitted to University of Nizwa

Student Paper

1%

3

Submitted to Texas A&M University -
Commerce

Student Paper

1%

4

Submitted to UC, San Diego

Student Paper

1%

5

export.arxiv.org

Internet Source

1%

6

Submitted to Higher Education Commission
Pakistan

Student Paper

1%

7

ixqwm.ballonskaufen.de

Internet Source

1%

Exclude quotes

Off

Exclude matches

Off

Deepfake proposal 1.1

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8