



Department of Computer Science
Gujarat University
5 Year Integrated M.Sc.(Computer Science)
Semester - VI

**Identification and Addressing Gender Bias in
Content**

Presented By:
Ansh Yadav (36)

Under the Guidance of :
Dr. Hardik Joshi



Index

Sr. No	Index	Page No.
1.	Project Profile.....	1
2.	Introduction.....	5
3.	Information Repository.....	6
4.	Data Mining.....	9
5.	Research Problem.....	11
6.	Methodology.....	12
7.	Literature Review.....	13
8.	System Architecture.....	17
9.	Data Dictionary.....	23
10.	Screenshots.....	25
11.	Future Work.....	37
12.	Bibliography.....	38



Project Profile

TITLE	DESCRIPTION
Project Title	Identification and Addressing Gender Bias in Content
Aim of the Project	Identify the gender biasness in content and categorize it into different category and sub categories
Project duration	6 month
Team size	1 Member
Tools used	Visual Studio code, Git hub
Technologies used	Python, Next.js, Flask, Tailwindcss, MongoDB
Guide	Dr. Hardik Joshi

Project Description



Introduction

In today's digital world, We see different content that plays an important role in changing our perception and behaviour toward the person or community. However, There exists a vast quantity of available content that may promote gender Bias.

As developers, we understand the importance of creating an unbiased content. It is crucial to identify and address the gender bias on the software. Identifying and addressing gender bias help the companies and business to track and stop the promotion of that bias in the business, help to safeguard the gender equality and create a safe environment in digital world



Information Repository

- 1) "id": Unique identifier for the tweet.
- 2) "tweet": Text of the tweet.
- 3) "labels_task1" : Indicates that tweet contains sexist behaviours
Values are: "YES" and "NO".
- 4) "labels_task2" : Indicates the intention of the person and categorize the tweet.
Values are: "DIRECT", "REPORTED", "JUDGEMENTAL".
- 5) "labels_task3" : Indicating the type of sexism that are found in the tweet.
Values are: "IDEOLOGICAL-INEQUALITY", "STEREOTYPING-DOMINANCE",
"OBJECTIFICATION", "SEXUAL-VIOLENCE", "MISOGYNY-NON-SEXUAL-VIOLENCE",.



Sample Dataset

```
{  
  "id": "200001",  
  "tweet": "Writing a uni essay in my local pub with a coffee. Random old man keeps asking me  
drunk questions when I'm trying to concentrate & ends with \"good luck, but you'll  
just end up getting married and not use it anyway\". #EverydaySexism is alive and well “,  
  "labels_task1": "['YES', 'YES', 'NO', 'NO', 'YES', 'NO']",  
  "labels_task2": "['JUDGEMENTAL', 'JUDGEMENTAL', '-', '-', 'REPORTED', '-']",  
  "labels_task3": "(['MISOGYNY-NON-SEXUAL-VIOLENCE'], ['SEXUAL-VIOLENCE'], ['-'], ['-  
'], ['MISOGYNY-NON-SEXUAL-VIOLENCE'], ['-'])"  
},
```



Data mining

- Check is it a unbiased or bias text
- Remove html tags
- Remove accents
- Spell checking
- Remove hastags
- Remove URL
- Remove punctuation
- Remove Number
- Remove Non grammatic word
- Lemmatize text



Data mining

- Extract only:
 - Noun (NN)
 - Verb, base form (VB)
 - Verb, past tense (VBD)
 - Verb, present participle (VBG)
 - Verb, Past Participle (VBN)
 - Verb, non-3rd person singular present (VBP)
 - Verb, 3rd person singular present (VBZ)



Sample Training Set

- Split in Chunk
- Category for 1 task
- Category for 2 task
- Sample Training Set :

```
{  
  "keyword": "novel",  
  "category_1": ["JUDGEMENTAL"],  
  "category_2": ["MISOGYNY-NON-SEXUAL-VIOLENCE"]  
},
```



Research Problem

- What are the methods and technique required to effectively identify and address gender bias in the content of software development and business ?
- The aim for Identification and Addressing Gender Bias in the content of a software development and business to stop the unbiasedness in the content of software development and promote safe environment in the digital realm and promote the unbiased and stop discrimination in the software



Methodology

1) Introduction:

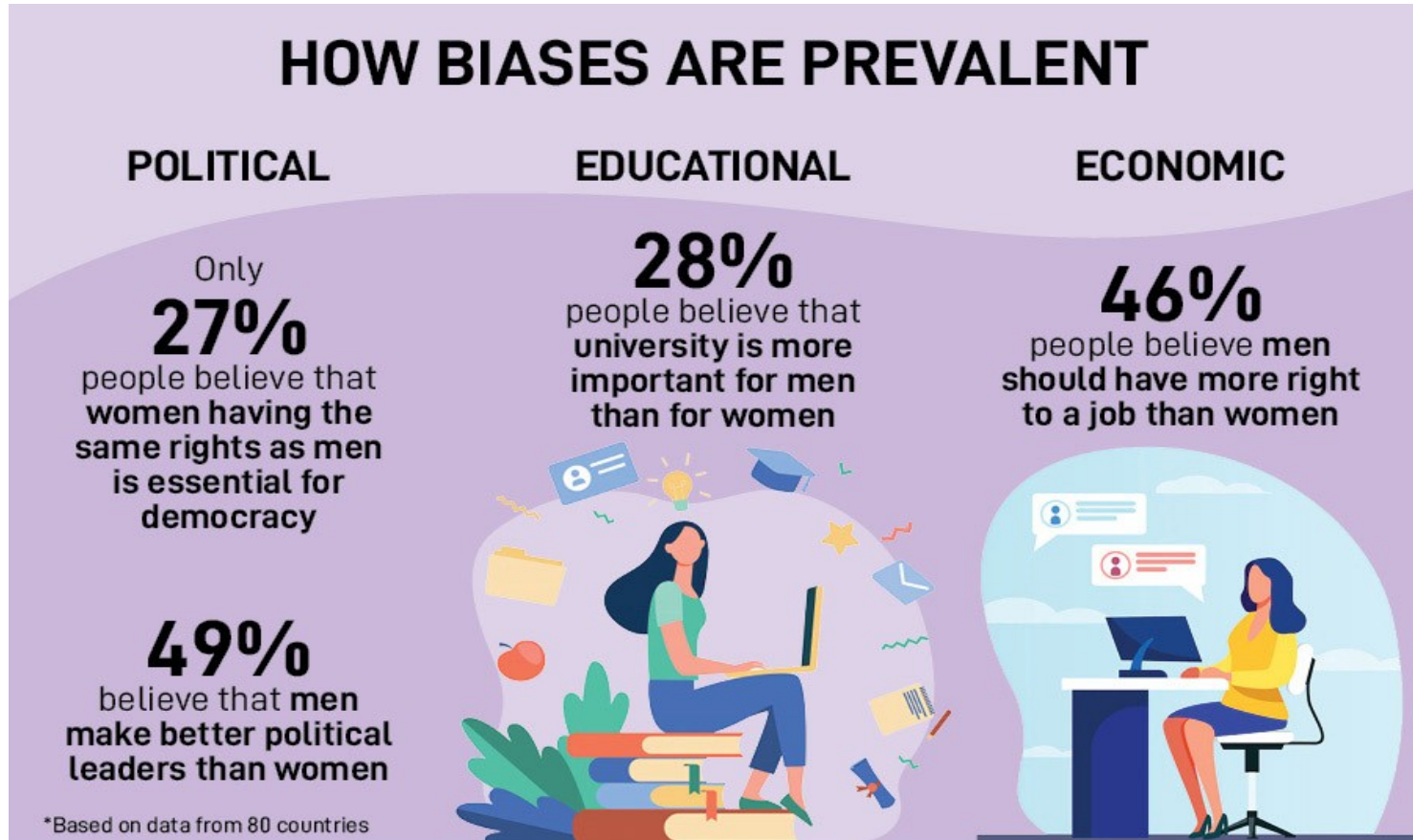
- Prompted the need for comprehensive analysis and mitigation strategies
- Aims to investigate the gender bias within the **EXIST 2024 dataset**

2) Research Design:

- Utilizing a mixed-methods approach
- research employs quantitative analysis to identify patterns
- providing deeper insights into the nature and implications of gender bias
- Aims to offer actionable recommendations for addressing gender bias



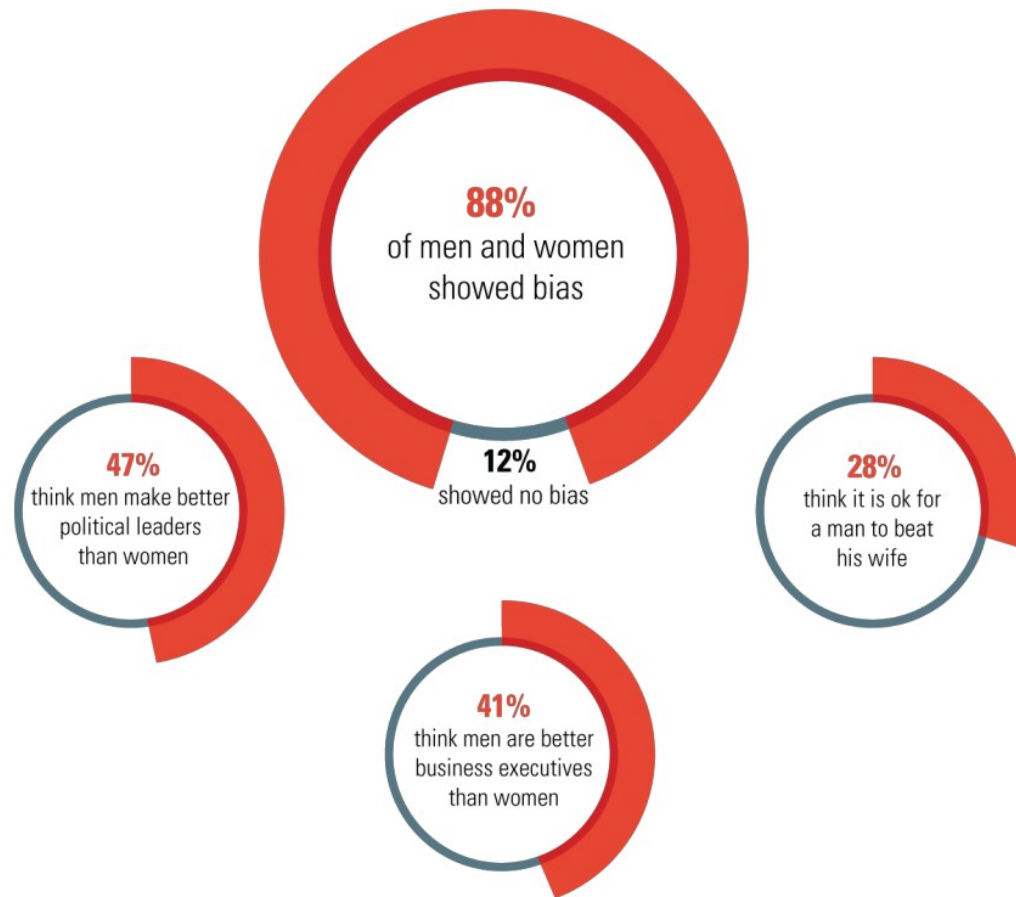
Literature Review (GSNI)



1. Gender Social Norm Index (GSNI) - 2020



Literature Review

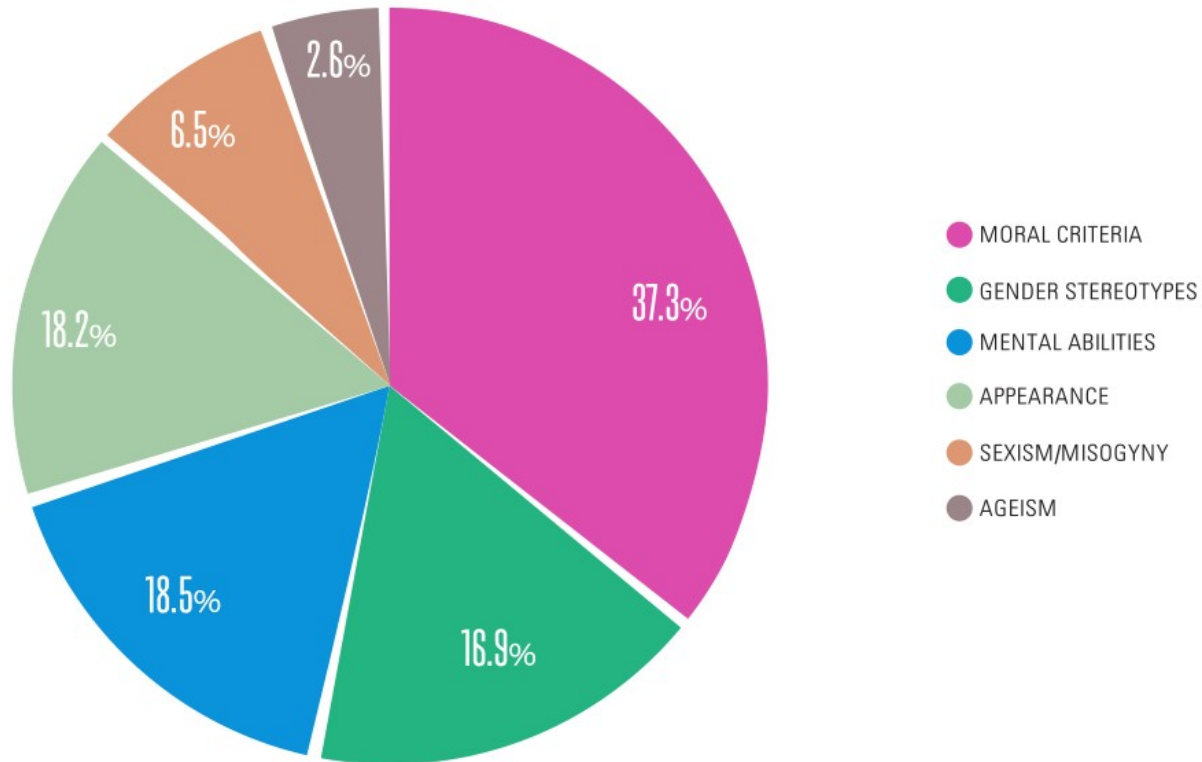


2. United Nation Development Programme (UNDP) - 2020



Literature Review

CHART 2. TYPOLOGY OF VIOLATIONS (%)

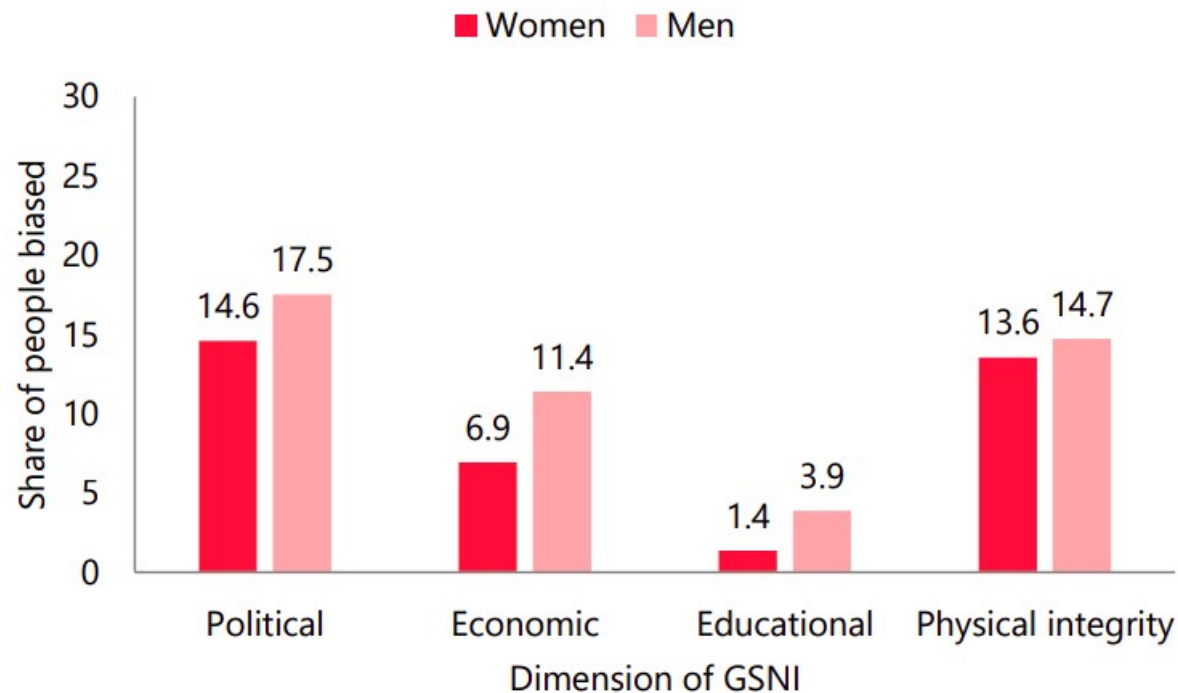


3. Media Development Foundation (MDF)- 2020



Literature Review

Share of people biased by sex for each dimension of the GSNI, last available period

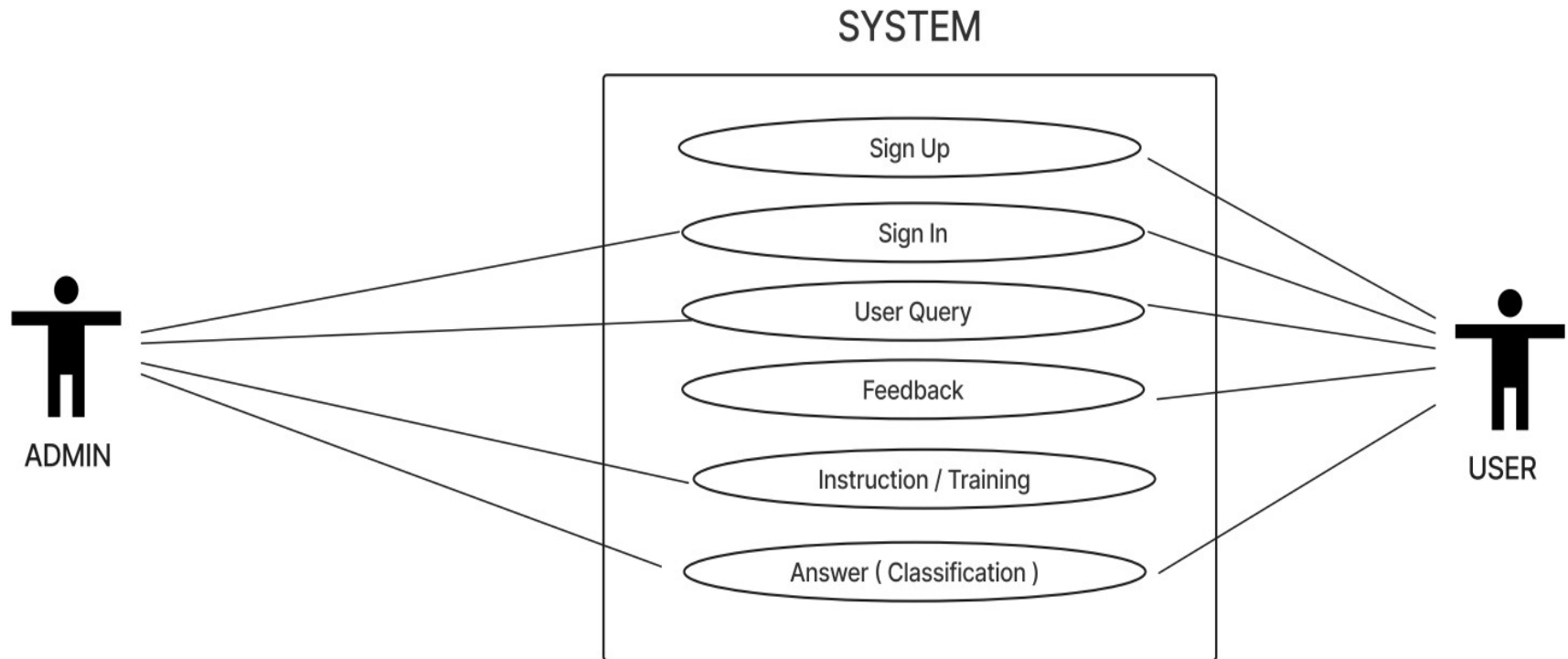


4. Sweden, GSNI - 2020

System Architecture

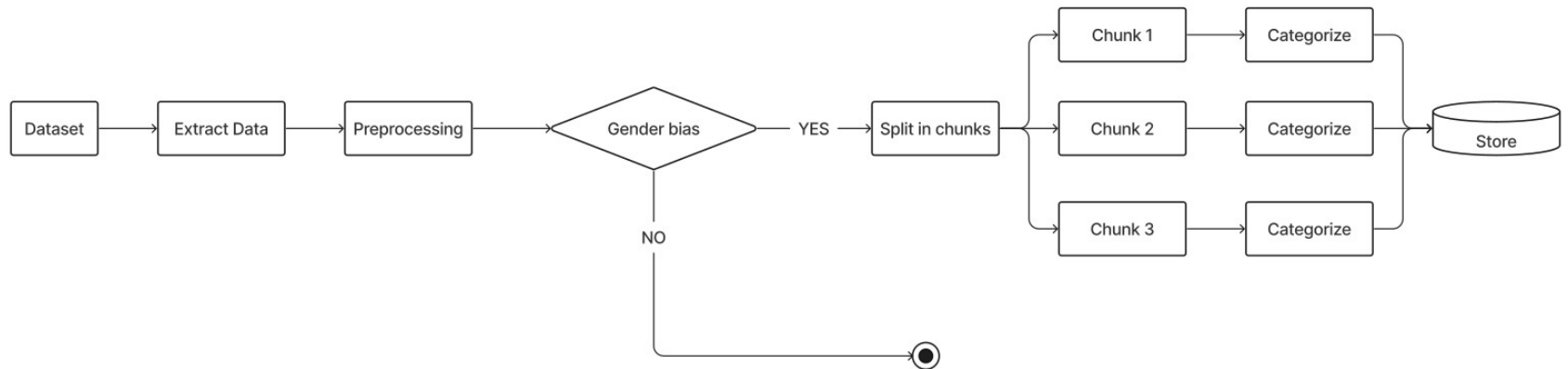


Use Case Diagram



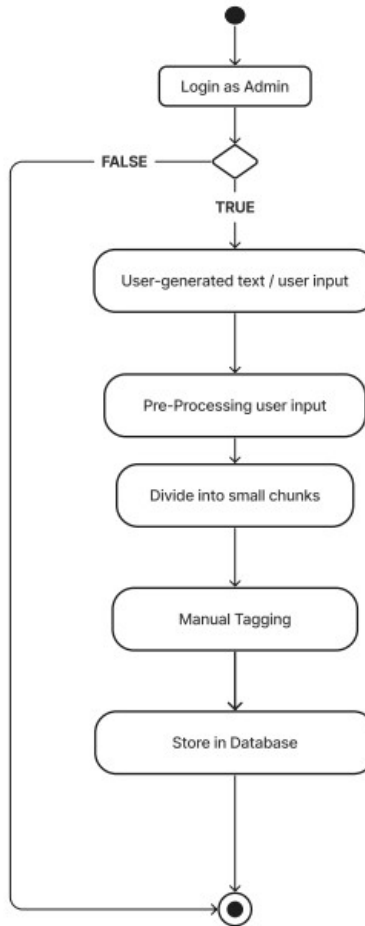


Process diagram



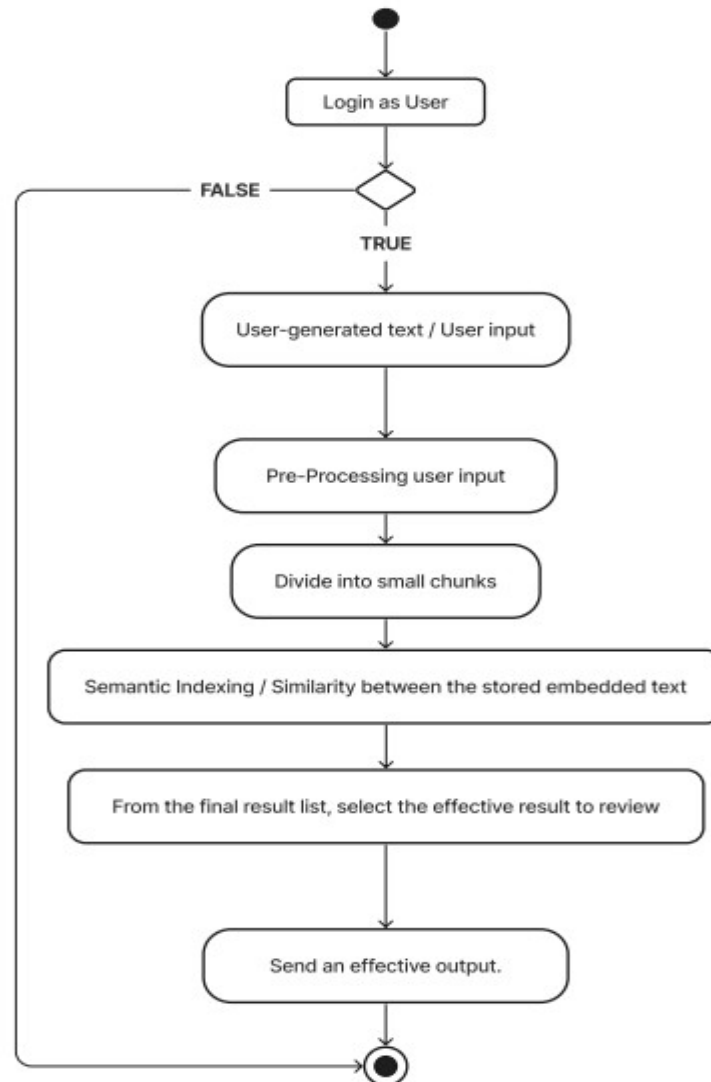


Activity Diagram of Training



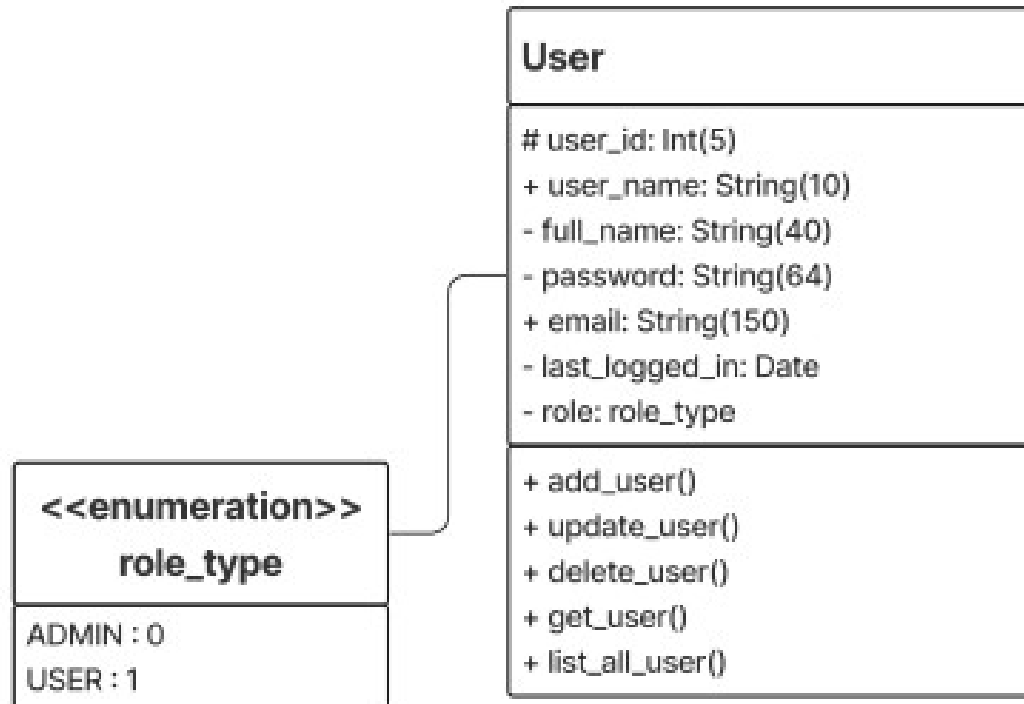


Activity Diagram of Query





Class Diagram



DATA DICTIONARY



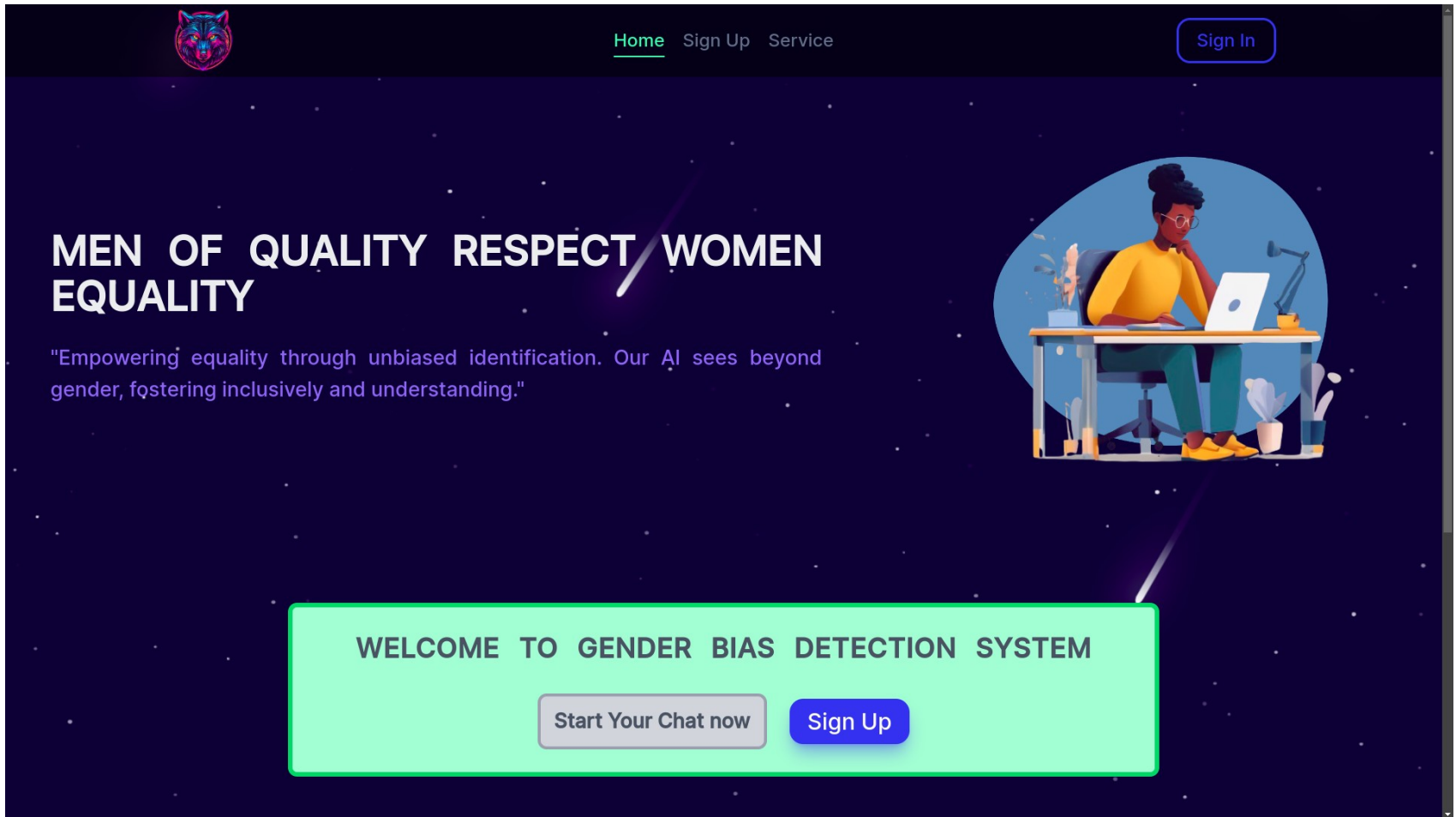
User Table

Seral no	Field name	Type	Size	Link to	Sample data
1	user_id	INT	24		65eabc234f7798 607df05963
2	user_name	STRING	10		John Wick
3	full_name	STRING	40	-	John Albert Warden Wick
4	password	STRING	64	-	952316575f25sd fe2a9cb5d035fe3 adsc72ddsdd
5	email	STRING	150	-	90901@gujunive rsity.ac.in
6	last_logged_in	DATE		-	2024-03-30
7	logged_in	DATE		-	2024-4-12
8	role_type	ENUM	0, 1		0

SCREENSHOTS

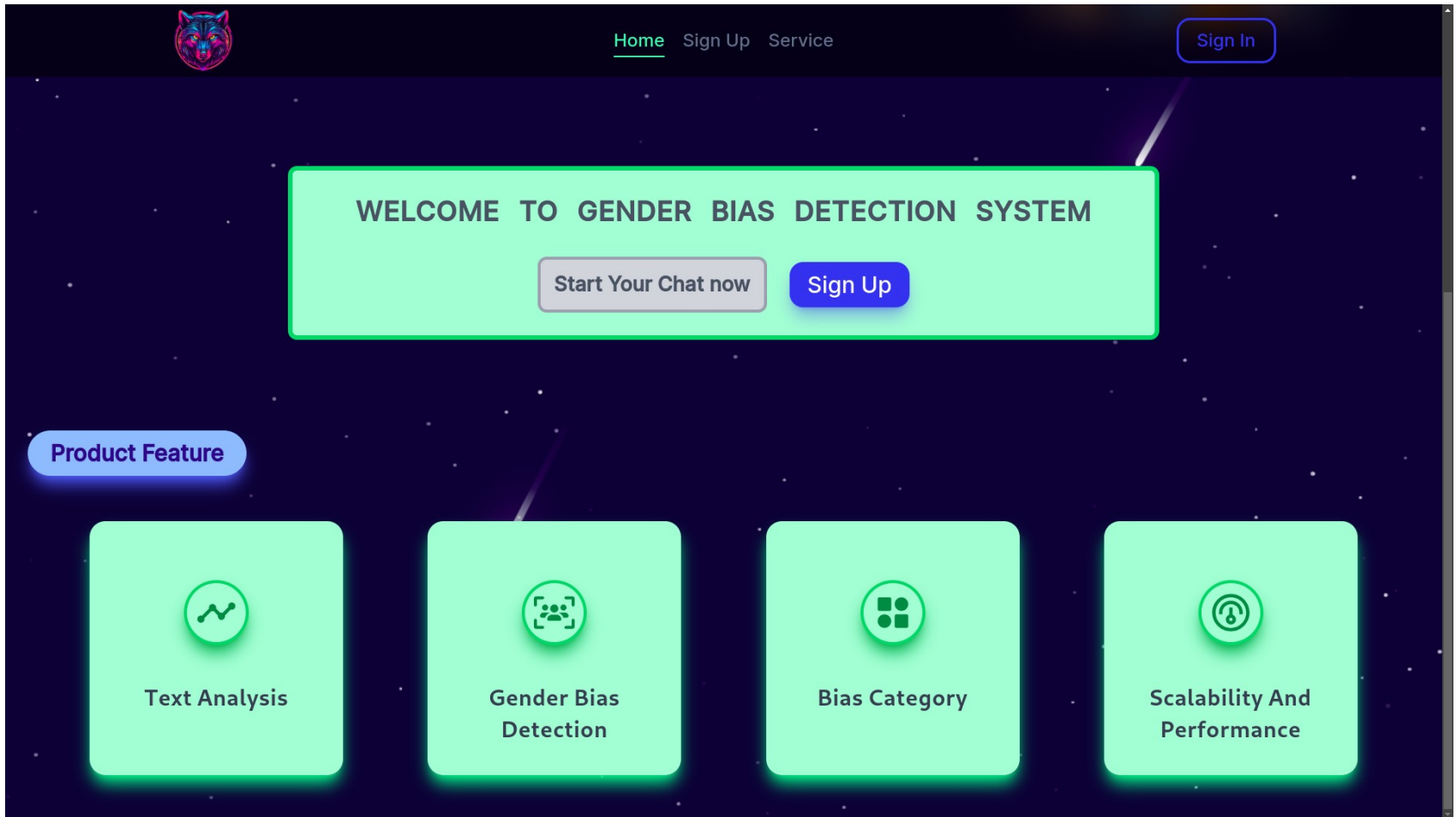


Home Page



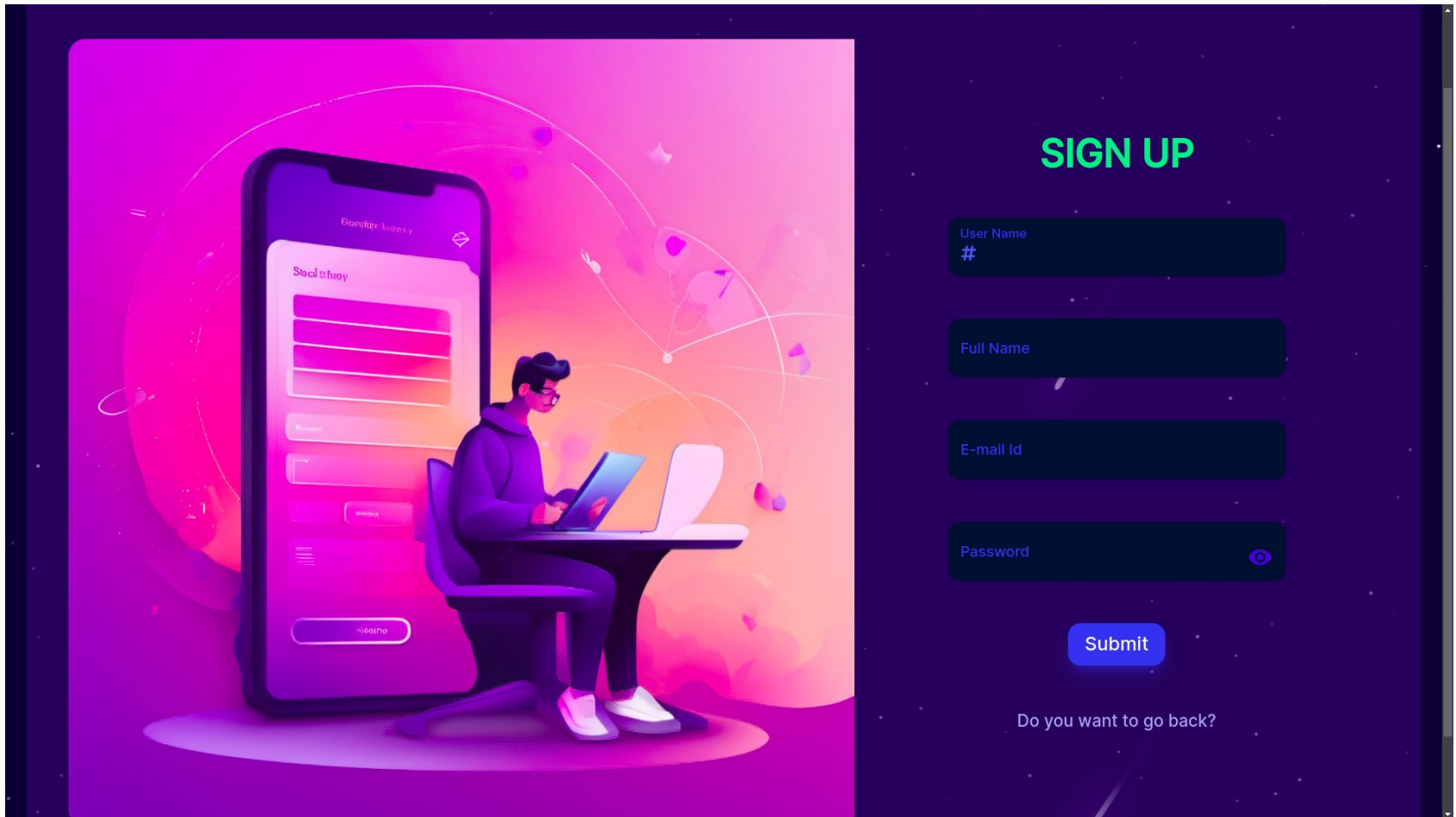


Home Page





Sign Up Page





Sign In Page

Home Sign Up Service Sign In

MEN OF QUALITY RESPECT WOMEN EQUALITY

"Empowering equality through unbiased identification, fostering inclusively and understanding."

LOG IN

User Name / E-mail
aj045045@gmail.com

Password
.....

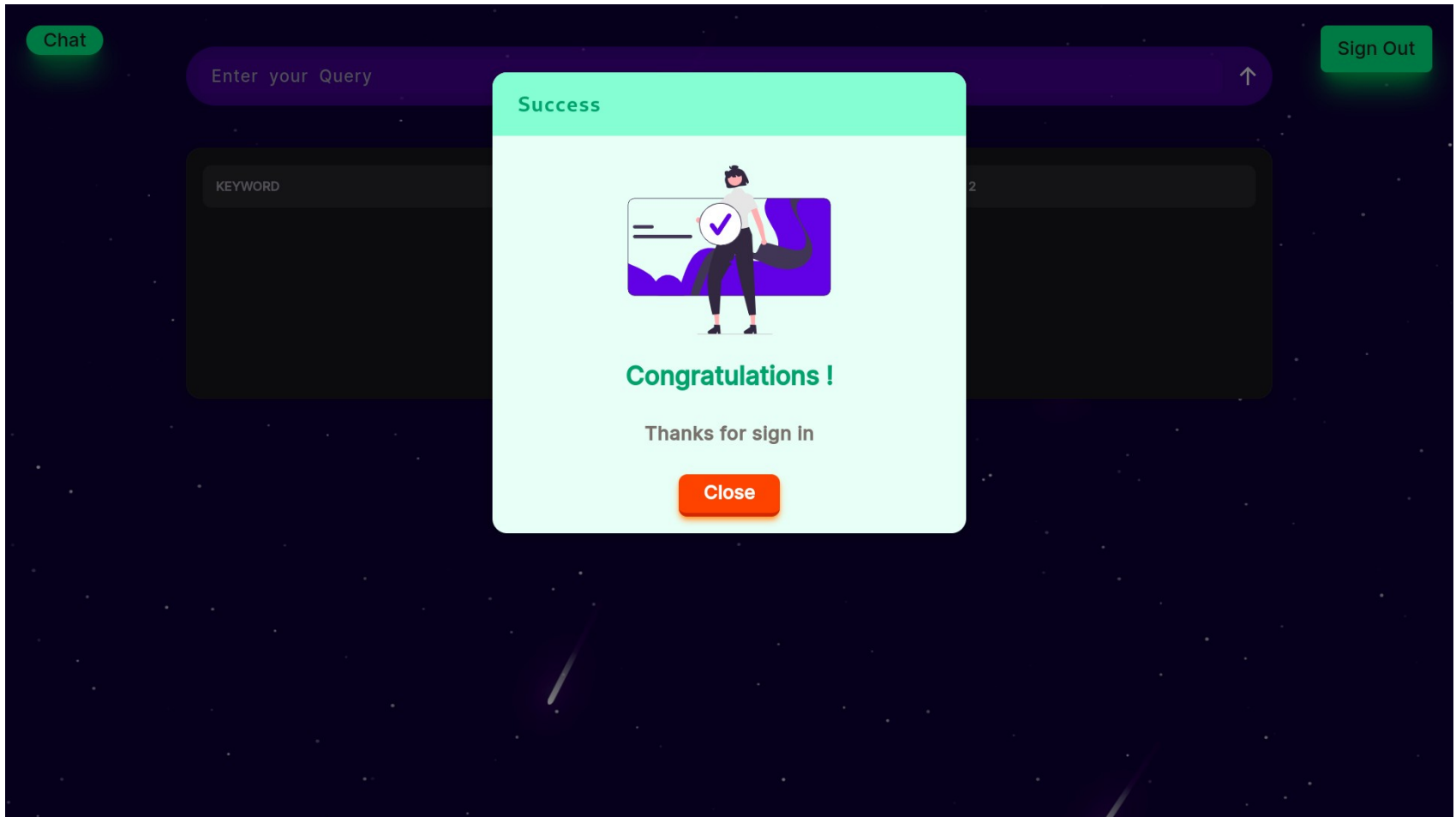
Close Sign in

WELCOME TO GENDER BIAS DETECTION SYSTEM

Start Your Chat now Sign Up

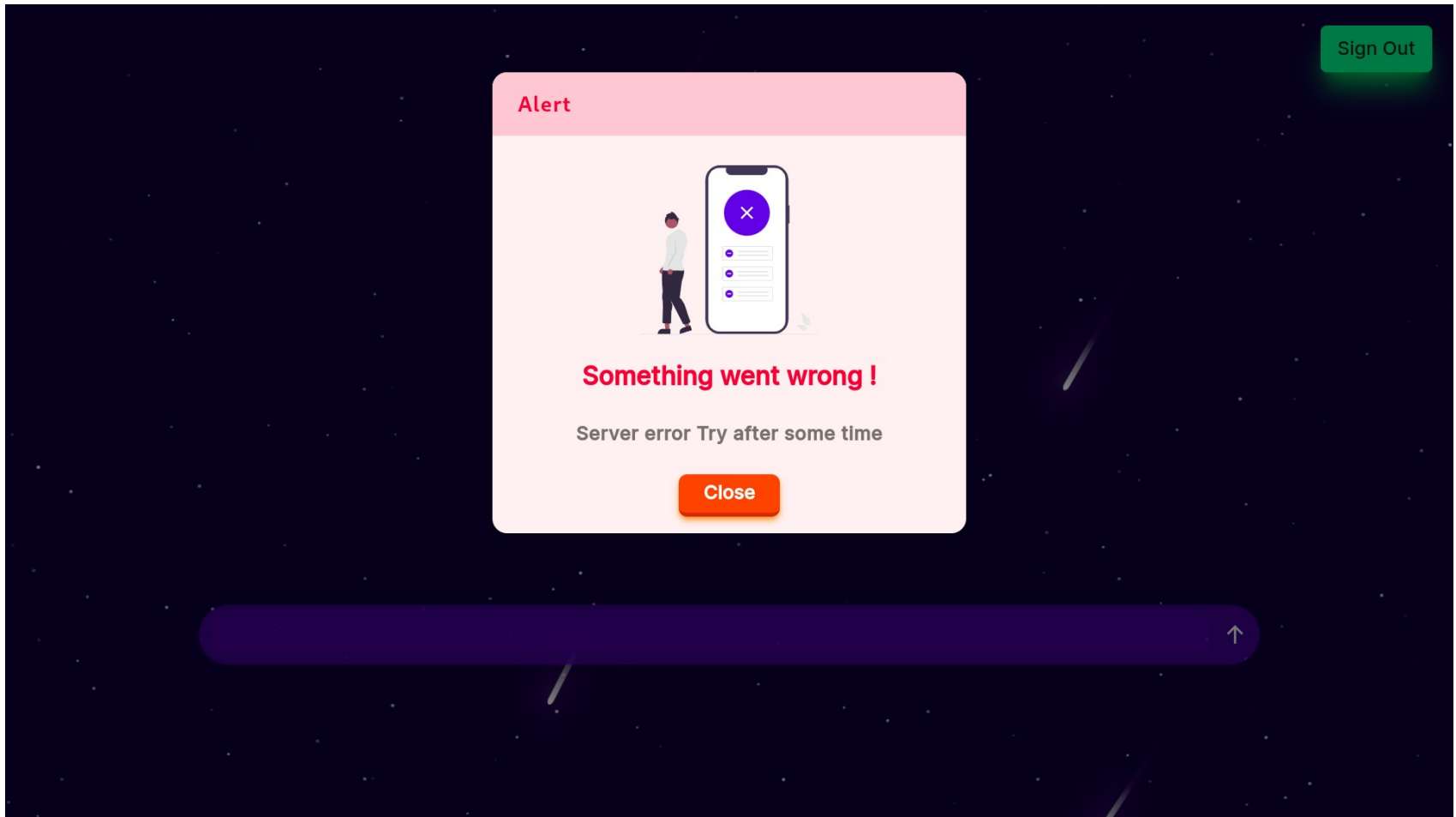


Success Message





Error Message





Admin Training Page

[Chat](#)

[Sign Out](#)

KEYWORD	CATEGORY 1	CATEGORY 2
Business	<input type="checkbox"/> Direct	<input type="checkbox"/> IDEOLOGICAL-INEQUALITY
	<input type="checkbox"/> Reported	<input type="checkbox"/> STEREOTYPING-DOMINANCE
	<input type="checkbox"/> Judgemental	<input type="checkbox"/> OBJECTIFICATION
		<input type="checkbox"/> SEXUAL-VIOLENCE
Owner		<input type="checkbox"/> MISOGYNY-NON-SEXUAL-VIOLENCE
	<input type="checkbox"/> Direct	<input type="checkbox"/> IDEOLOGICAL-INEQUALITY
	<input type="checkbox"/> Reported	<input type="checkbox"/> STEREOTYPING-DOMINANCE
	<input type="checkbox"/> Judgemental	<input type="checkbox"/> OBJECTIFICATION
Satisfaction		<input type="checkbox"/> SEXUAL-VIOLENCE
	<input type="checkbox"/> Direct	<input type="checkbox"/> MISOGYNY-NON-SEXUAL-VIOLENCE
	<input type="checkbox"/> Reported	<input type="checkbox"/> IDEOLOGICAL-INEQUALITY
	<input type="checkbox"/> Judgemental	<input type="checkbox"/> STEREOTYPING-DOMINANCE



Admin Training Page

[Chat](#)

Product

☐ Direct
☐ Reported
☐ Judgemental

☐ IDEOLOGICAL-INEQUALITY
☐ STEREOTYPING-DOMINANCE
☐ OBJECTIFICATION
☐ SEXUAL-VIOLENCE
☐ MISOGYNY-NON-SEXUAL-VIOLENCE

Service

☐ Direct
☐ Reported
☐ Judgemental

☐ IDEOLOGICAL-INEQUALITY
☐ STEREOTYPING-DOMINANCE
☐ OBJECTIFICATION
☐ SEXUAL-VIOLENCE
☐ MISOGYNY-NON-SEXUAL-VIOLENCE

Need

☐ Direct
☐ Reported
☐ Judgemental

☐ IDEOLOGICAL-INEQUALITY
☐ STEREOTYPING-DOMINANCE
☐ OBJECTIFICATION
☐ SEXUAL-VIOLENCE
☐ MISOGYNY-NON-SEXUAL-VIOLENCE

Choice

☐ Direct
☐ Reported
☐ Judgemental

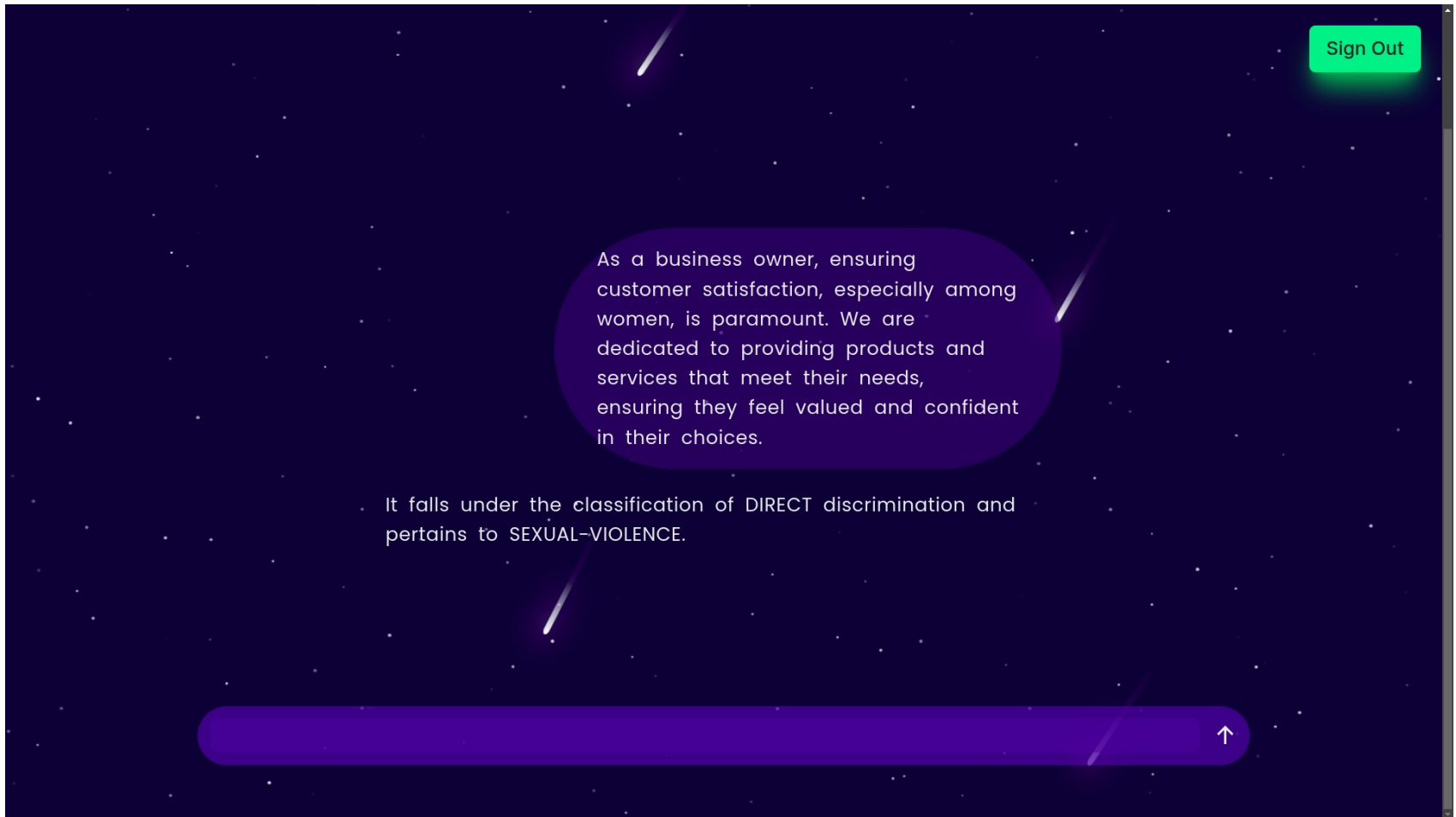
☐ IDEOLOGICAL-INEQUALITY
☐ STEREOTYPING-DOMINANCE
☐ OBJECTIFICATION
☐ SEXUAL-VIOLENCE
☐ MISOGYNY-NON-SEXUAL-VIOLENCE

[Sign Out](#)

[Add Keywords](#)

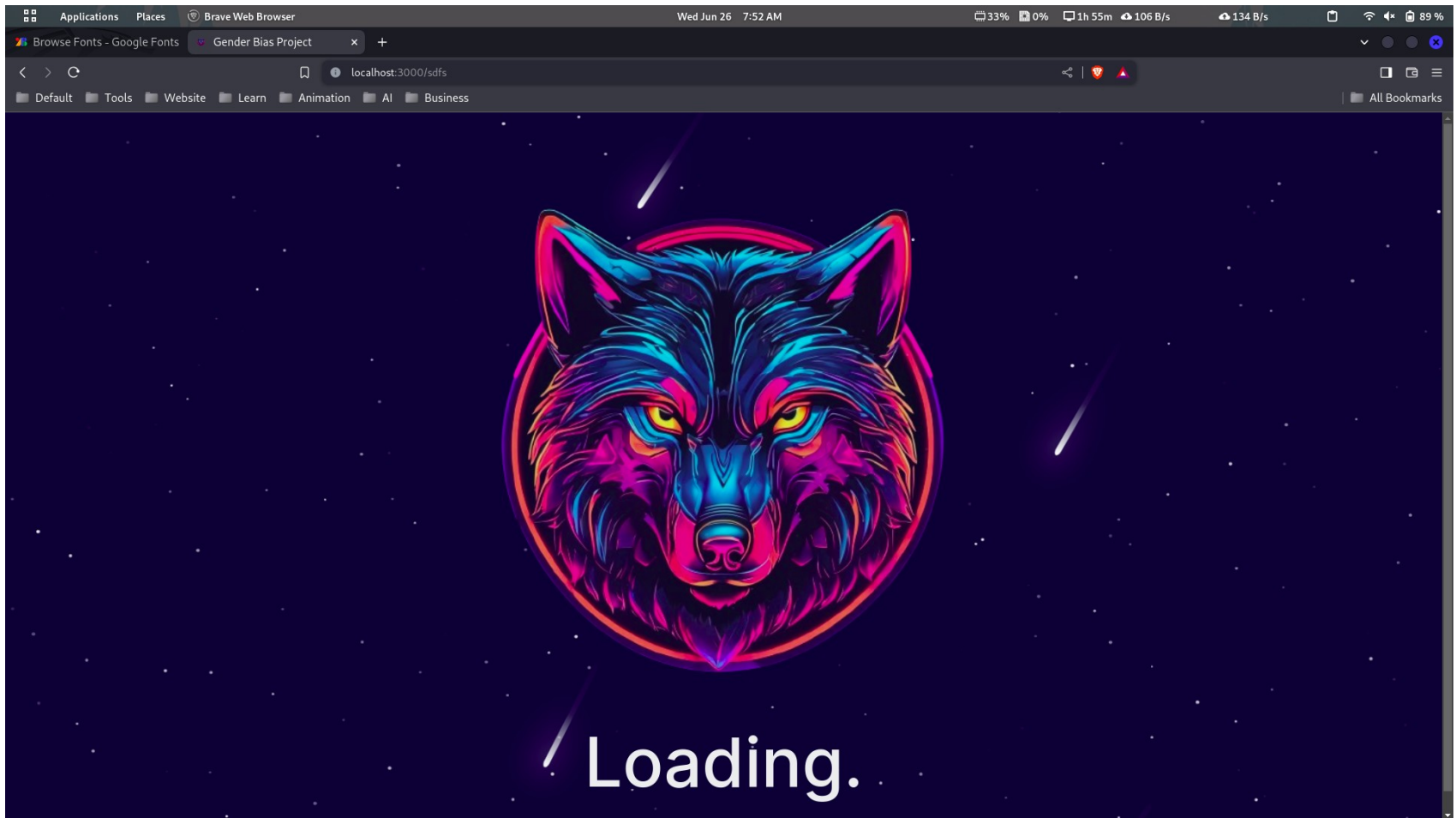


Chat Page





Loading Page





Error Page





Future Work

- Integrate meme / Image recognition
- Audio file processing
- Implementing File Reading
- Apply machine learning technique



Bibliography

- **Googlescholar** :https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=gender+bias+in+media&btnG=&oq=gender
- **Kaggle:** <https://www.kaggle.com/datasets/thedevastator/winobias-a-winograd-schema-dataset-for-gender-bi>
- **Github:** https://github.com/yandexdataschool/nlp_course
- **Flask-** <https://flask.palletsprojects.com/en/3.0.x/>
- **Redux -** <https://redux.js.org/>

THANK YOU