**Mohamed Khider University of Biskra**
Faculty of Exact Sciences
Department of Computer Science

# MASTER THESIS

Speciality : Master - Artificial Intelligence - Ref : P1065/2026

TITLE

## An LLM-Based Communication Assistant for Deaf and Non-Verbal Individuals

Prepared and defended by

**Elmouataz billah NOUAR**

**Mohamed lokrichi NID**

Defense date : 08/06/2026

DEFENSE COMMITTEE

| Full Name | Grade | Role |
|---|---|---|
| **Somia SAHRAOUI** | Prof | **President** |
| **Zohra HMIDI** | MCB | **Supervisor** |
| **Mohamed Nadjib MEADI** | MCA | **Examiner** |
| / | - | **Incubator Representative** |

**Academic Year : 2025/2026**

# Table Of Contents

List Of Contents I
List Of Tables II
List Of Figures III
Abstract IV
General Introduction 1
**Chapter1- Evolution of Sign Language ..... 3**
1.1 Introduction ..... 4
1.2 Overview of Sign Language ..... 4
1.3 Sign Langage ..... 5
1.3.1 Definition ..... 5
1.3.2 Types of Sign Language ..... 5
1.3.2.1 Natural Sign Languages ..... 5
1.3.2.2 Village Sign Languages ..... 5
1.3.2.3 International Sign (IS) ..... 6
1.3.2.4 Tactile Sign Languages (for DeafBlind people) ..... 6
1.3.2.5 Finger-spelling/ Manual Alphabets ..... 8
1.3.3 Uses of sign language ..... 9
1.3.4 Communication Problems in Sign Language ..... 9
1.4 Existing Sign Language Databases ..... 10
1.4.1 Isolated Sign Language Recognition Dataset (ISLRD) ..... 10
1.4.2 Continuous Sign Language Recognition Dataset (CSLRD) ..... 10
1.4.3 Sign Language Translation Dataset (SLTD) ..... 11
1.4.4 Finger-spelling / Alphabet Recognition Dataset ..... 11
1.4.5 Multi-Modal / 3D Sign Recognition Dataset ..... 12
1.5 Existing and Domains of Assistive Applications ..... 12
1.5.1 Hand Talk ..... 12
1.5.2 Lingvano ..... 13
1.5.3 ASL Dictionary (Software Studios) ..... 14
1.5.4 Sorenson ntouch (VRS) ..... 14
1.6 Necessity of Implementing Recent Applications ..... 15
1.7 Conclusion ..... 16
**Chapter2- Learning Techniques ..... 17**
2.1 Introduction ..... 18
2.2 Model-based Training Techniques ..... 18
2.2.1 Machine Learning Techniques ..... 18

2.2.1.1 Support Vector Machines (SVM) ..... 18
2.2.1.2 Multi-Layer Perceptron (MLP) ..... 19
2.2.1.3 Decision Trees / Random Forests ..... 20
2.2.2 Deep Learning Techniques ..... 22
2.2.2.1 Convolutional Neural Network(CNN) ..... 22
2.2.2.2 Recurrent Neural Networks (RNNs) ..... 23
2.2.2.3 Long Short-Term Memory (LSTM) ..... 24
2.3 AI Paradigm-based Techniques ..... 25
2.3.1 Large Language Models (LLMs) ..... 25
2.3.2 Transformer ..... 27
2.3.3 Computer Vision (CV) ..... 30
2.4 Related Works ..... 31
2.5 Conclusion ..... 32
**Chapter3- Application Design ..... 33**
3.1 Introduction ..... 34
3.2 Data Description ..... 34
3.2.1 Data Pre-processing ..... 35
3.3 General architecture of the application ..... 37
3.3.1 Model Pipeline ..... 37
3.3.2 Communication Modes ..... 39
3.3.2.1 Communication Mode 1 ..... 39
3.3.2.2 Communication Mode 2 ..... 40
3.3.3 Peer-to-Peer Communication ..... 41
3.3.4 User Interface ..... 42
3.4 System Design With Uml ..... 42
3.4.1 UML ..... 42
3.4.2 Use Case diagrams ..... 43
3.4.3 Class diagrams ..... 43
3.4.4 Sequence diagrams ..... 44
3.4.5 Activity diagrams ..... 45
3.5 Database design ..... 46
3.5.1 Association entity model ..... 47
3.5.2 Entity–relationship model ..... 47
3.6 Conclustion ..... 47

**Chapter4- Implementation and Results** ..... **48**
4.1 Introduction ..... 49
4.2 Development environment ..... 49
4.2.1 Programming language ..... 49
4.2.1.1 Python ..... 49
4.2.1.2 HTML(Hyper Text Markup Language) ..... 50
4.2.1.3 CSS(Cascading Style Sheets) ..... 50
4.2.1.4 JavaScript ..... 50
4.2.1.5 MySQL ..... 50
4.2.2 Libraries ..... 50
4.2.2.1 Pyhorch ..... 51
4.2.2.2 MediaPipe ..... 51
4.2.2.3 OpenCV ..... 51
4.2.2.4 Flask ..... 51
4.2.2.5 MoviePy ..... 51
4.2.2.6 Cryptography ..... 51
4.2.2.7 BcryptJS ..... 51
4.2.2.8 Socket (Python) ..... 52
4.3 Model Performance and Analysis ..... 52
4.3.1 Training Details ..... 52
4.3.2 Model Results and Analysis ..... 52
4.3.3 Result discussion ..... 55
4.3.3.1 Result evaluation ..... 55
4.3.3.2 Result comparison ..... 56
4.4 Overview of Application Interface ..... 57
4.4.1 Login ..... 57
4.4.2 Sign-Up ..... 58
4.4.3 Dashboard ..... 59
4.4.4 User settings ..... 60
4.4.5 Normal person chat ..... 61
4.4.6 Deaf person chat ..... 63
4.5 Conclusion ..... 64

**General conclusion** **65**
**Bibliography** **66**

# List of Tables

2.1 Summary of Sign Language Detection Papers Using LSTM . . . . . 31
4.1 Characteristics of the used material . . . . . 49
4.2 Evaluation results of classification bilstm model . . . . . 53

# List of Figures

1.1 Tactile Sign Languages:TASL (a) and TBSL (b) . . . . . 7
1.2 Manual Alphabets: ASL (a) and BSL (b) . . . . . 8
1.3 Hand Talk Application . . . . . 12
1.4 Lingvano Application . . . . . 13
1.5 ASL Dictionary Application . . . . . 14
1.6 Sorenson ntouch Application . . . . . 14
2.1 SVM Architecture . . . . . 19
2.2 MLP Architecture . . . . . 20
2.3 DT and RF Architecture . . . . . 21
2.4 CNN Architecture . . . . . 22
2.5 RNN Architecture . . . . . 23
2.6 LSTM Architecture . . . . . 24
2.7 LLM Architecture . . . . . 26
2.8 Transforme Architecture . . . . . 29
2.9 Computer Vision Architecture . . . . . 30
3.1 Properties of existing ISLR datasets for ASL compared to ASL Citizen . . . 35
3.2 Sign Language Data Pre-processing Workflow . . . . . 36
3.3 General architecture of the application . . . . . 37
3.4 Normal person use case diagram . . . . . 43
3.5 Deaf person use case diagram . . . . . 43
3.6 Class diagram . . . . . 44
3.7 User sequence diagram . . . . . 45

3.8 User activity diagram . . . . . 46
3.9 Database association entity model . . . . . 47
4.1 Accuracy and training loss Curve of BiLSTM model . . . . . 54
4.2 Confusion Matrix of BiLSTM model . . . . . 55
4.3 Login page view . . . . . 57
4.4 Sign-Up page view . . . . . 58
4.5 Dashboard page view . . . . . 59
4.6 User Settings page view. . . . . 60
4.7 Normal person chat page view. . . . . 61
4.8 User Reporting view. . . . . 62
4.9 Deaf person chat page view. . . . . 63
4.10 Deaf person sign recognition view. . . . . 64

## Acknowledgment

*Praise be to God Almighty, who taught us what we did not know. After that, I can only extend my sincere and special thanks to Dr. Hmidi Zahra, who kindly supervised our project, read the report, corrected it, removed the confusion, and helped us complete our academic career and crown it with this blessed scientific fruit. I also thank my family for supporting me*

# Abstract

The aim of this project is to apply computer vision and AI techniques in order to build a bi-directional intelligent communication assistant that will make interactions between hearing and deaf-mute people easier and more effective. Using this application, it is possible to transform sign language gestures to a readable text or speech message, and vice versa in order to minimize the gap in communication between hearing-impaired people and those who can hear.

The architecture includes two main modules. In the first one, MediaPipe is used to recognize shoulder and hand landmarks based on the captured sign language gestures via a camera. After obtaining the landmarks, an LSTM model trained on the ASL citizen dataset is used in order to determine the corresponding word. Then, with the help of a big language model like Gemini, the obtained words are put together in a grammatically correct sentence. The second module operates the opposite way by accepting text messages provided by the user, tokenizing the sentence and finding the videos of the corresponding sign languages that are located within the dataset. The sequence of these videos is played one after another in order to translate the sentence in sign language.

This application is easy to use, accessible and highly interactive and thus it was built using web technologies, such as HTML, CSS, and JS for frontend, while Flask works as a backend in this case. Gemini's API is used for translation purposes, and PyTorch together with MediaPipe were applied in this project in order to detect gestures.

As a result of this project, it became possible to provide two-way communication and interaction between deaf and mute individuals and others.

# General Introduction

The foundation of human communication is language, which enables people to interact, exchange ideas, and fully engage in society. Sign languages offer deaf and hard-of-hearing individuals a rich visual-manual communication method that is just as linguistically valid as spoken languages. These languages convey meaning by combining body posture, facial expressions, hand shapes, and movement.

In recent years, the field of artificial intelligence (AI) has opened up new avenues for supporting sign language users. Through the use of computer vision, machine learning, and gesture recognition, an increasing number of artificial intelligence (AI) systems are being developed to recognize sign language, translate it into text or speech, and assist in removing communication barriers. The study of artificial intelligence (AI) and sign language together holds great promise for enhancing accessibility and encouraging the full involvement of deaf communities in everyday and online interactions.

## 1 Challenges and solutions

The most challenges faced in sign language recognition systems are:

1. **Lack of standardized datasets:** Many sign languages, especially local or regional ones such as Algerian Sign Language, suffer from the absence of large, well-annotated datasets, which limits model training and accuracy.
2. **Variation in gestures:** The same sign can differ in meaning or form across regions or even individuals, creating ambiguity and reducing recognition reliability.
3. **Complexity of motion and facial expressions:** Sign language involves hand movements, facial expressions, and body posture, making it difficult for models to capture all visual features effectively.
4. **Lack of technological accessibility:** In many regions, there is limited access to high-quality cameras or computing resources needed to develop and deploy sign language systems.
5. **Limited linguistic and cultural documentation:** The absence of a unified dictionary or linguistic framework for some sign languages, such as Algerian SL, complicates data collection and interpretation.

Despite this, researchers have tried to eliminate these problems by:

1. **Developing computer vision systems:** by using deep learning and neural networks to recognize hand gestures and movements accurately.
2. **Creating collaborative datasets:** through cooperation with deaf communities to collect and label reliable sign language data.

4. **Enhancing linguistic models:** integrating natural language processing (NLP) with gesture recognition to interpret meaning and context.
5. **Designing assistive applications:** building real-time translation and communication tools to bridge the gap between deaf and hearing individuals.
6. **Promoting accessibility and education:** encouraging public awareness and AI-based learning platforms for inclusive communication.

# 2 Work objective

According to the research, it was concluded that the most effective way to address these issues is to combine computer vision and artificial intelligence, which makes it possible to recognize and translate sign language . This method encourages accessibility in day-to-day interactions and aids in communication between hearing and deaf people.

The goal of our project, "An LLM-Based Communication Assistant for Deaf and Non-Verbal Individuals," is to create a tool that will help the deaf community communicate with the general public. In order to promote inclusivity, communication ease, and global understanding. It aims to offer an intelligent application that can translate gestures into text or speech in real time.

This work will consist of four chapters:

• Chapter 1: "Evolution of Sign Language ",This chapter introduces sign language by presenting its definition, different types, and areas of use. It also highlights the main communication challenges faced by deaf individuals. Furthermore, it reviews existing datasets and explores current applications developed to support this community, along with their various domains of use. Finally, the chapter discusses the limitations of these solutions and emphasizes the need for developing more advanced and modern applications.

• Chapter 2:"Learning Techniques",This chapter presents AI methods for sign language recognition, focusing on computer vision for feature extraction and Transformer-based models for sequence and context understanding. It also reviews related work, highlighting the strengths and limitations of existing approaches.

• Chapter 3: "Application design",This chapter presents the system design , including the database design, model architecture, and LSTM-based gesture recognition. It also illustrates the system through use case, class, sequence, and activity diagrams, along with the general architecture and associative entity model.

• Chapter 4: "Implementation and Discus of results", This chapter describes the actual implementation of the Sign Language Recognition (SLR) system, including the hardware and software used, the experimental configuration, the display and evaluation of model performance outcomes, and the user interface design.

At the end of this thesis, we arrive at a general conclusion that summarizes the main findings of the research and describes some future views.

# Chapter1- Evolution of Sign Language

## 1.1 Introduction

Sign language is not merely a language, it is a fundamental aspect of human expression, representing not only individuals but also a piece of the society in which they live.through gestures, expressions, and movements, it captures culture, identity, and the shared experiences of Deaf communities.understanding sign language, therefore, is not only about communication but also about acknowledging the social, cultural, and technological frameworks that support it.

This chapter provides a comprehensive exploration of the landscape surrounding sign language, its resources, and technological support systems.it begins by examining the concepts, types, and uses of sign language, as well as the challenges faced in communication, providing context for why research and innovation in this field are essential.following this, the chapter reviews existing sign language datasets, highlighting the variety, scope, and accessibility of resources available for both research and practical applications.

Building on this foundation, attention is given to current applications designed to assist Deaf and hard-of-hearing individuals, describing their functionalities, areas of implementation, and impact on daily life.finally, the chapter concludes with a discussion on recent advances and the ongoing need for innovative applications, emphasizing opportunities to bridge gaps in communication and enhance the effectiveness of sign language technologies.

## 1.2 Overview of Sign Language

Outside of its linguistic and cultural contexts, the term "sign language" is broad and frequently ill-defined. handshape, palm orientation, location (where the sign is articulated in space or on the body), movement (path and manner of motion), and non-manual features (such as facial expressions and body posture) are the five fundamental phonological parameters of a sign [1], though they are not the only ones[2]. signs are the vocabulary of a visual-spatial language controlled by these structured elements, similar to the vowels and consonants of spoken languages, they are not pantomime[3] .

The dimensions outlined and covered by sign language recognition requirements vary widely amongst literature and technological applications[4].

For example, **Stokoe** asserted in 1960 that sign languages possess highly developed grammatical systems, contradicting the notion that signs are merely gestures or silent representations[5]. **Klima** and **Beloge** also emphasized the importance of non-manual features in 1979, arguing that facial expressions and body movements are essential for conveying grammatical and semantic information[6].furthermore, **Becker-Schenck** and **Cockley** argued in 1980 that understanding the phonetic structure of signs is crucial for linguistic analysis and effective communication, as subtle changes in hand shape or movement can significantly alter meaning[7]. These findings demonstrate that sign languages are complex and structured systems that require the same level of analytical scrutiny as spoken languages.

# 1.3 Sign Langage

## 1.3.1 Definition

Sign language is a natural, fully developed language that uses visual-manual modality rather than the auditory-vocal modality. It consists of hand shapes, movements, orientations, locations, facial expressions, and body postures to convey meaning, grammar, and syntax. Sign languages are not universal; each country or community may have its own distinct sign language, complete with its own rules and vocabulary[5].

## 1.3.2 Types of Sign Language

### 1.3.2.1 Natural Sign Languages

- **Definition**

  According to Stokoe, William C. (1960) in "Sign Language Structure", natural sign languages are fully developed languages that arise naturally within Deaf communities. They have their own phonology, morphology, syntax, and semantics — independent of the spoken languages around them[5].

- **Examples**

  1. **American Sign Language (ASL):** A natural sign language used primarily in the USA and parts of Canada, with its own grammar and vocabulary, independent of English.
  2. **British Sign Language (BSL):** Used in the United Kingdom, BSL is a fully developed language with unique syntax and expressions, distinct from ASL and spoken English.
  3. **French Sign Language (LSF):** The primary sign language in France, LSF has its own linguistic structure and is historically significant, influencing many other sign languages worldwide.
  4. **Chinese Sign Language (CSL):** Used in China, CSL has regional variations and its own grammar, serving as the main language of communication for the Deaf community there.
  5. **Algerian Sign Language (LSA):** Used in Algeria, LSA is a natural sign language developed by the Algerian Deaf community, with its own vocabulary and grammar, influenced by French Sign Language.

### 1.3.2.2 Village Sign Languages

- **Definition**

  Zeeshan Ulrich and Connie de Vos (2012) explained in their book "Sign Languages in Village Communities: Anthropological and Linguistic Insights" that rural sign languages are sign languages that arise naturally within small and often isolated communities where hereditary deafness is common, prompting them to use signs daily[8].

- **Examples**
  1. **Kata Kolok (Bali, Indonesia):** Used in Bengkala, Bali, where hereditary deafness has been present for generations. Both Deaf and hearing villagers communicate fluently in it. The language is unique in its structure and does not borrow from Indonesian Sign Language. It features distinctive gestures such as foot-stamping to attract attention and makes rich use of spatial references[9].
  2. **Adamorobe Sign Language (Ghana):** Used in the village of Adamorobe in Ghana, where a high incidence of hereditary deafness led to its development. It is linguistically independent from Ghanaian Sign Language and is characterized by a simplified grammatical structure, lacking features like verb agreement[10].

#### 1.3.2.3 International Sign (IS)

- **Definition**

  As Rosenstock and Napier (2016) pointed out in "International Sign Language: Linguistic, Usage and Status Issues," international sign language is a type of communication used in international assemblies of the deaf that combines widely known signs and gestures[11].

- **Examples**
  1. **World Federation of the Deaf (WFD) Meetings:**IS is used during official WFD gatherings to allow delegates from different countries to communicate without relying on spoken or written translations.
  2. **Deaflympics (International Deaf Sports Events):**Athletes and organizers from multiple countries use IS for announcements, meetings, and social interactions. The language ensures inclusivity when participants speak different national sign languages.
  3. **International Conferences and Workshops:**IS is often employed in academic or advocacy settings, such as workshops on Deaf rights or linguistic research, to facilitate understanding among Deaf scholars from various countries.

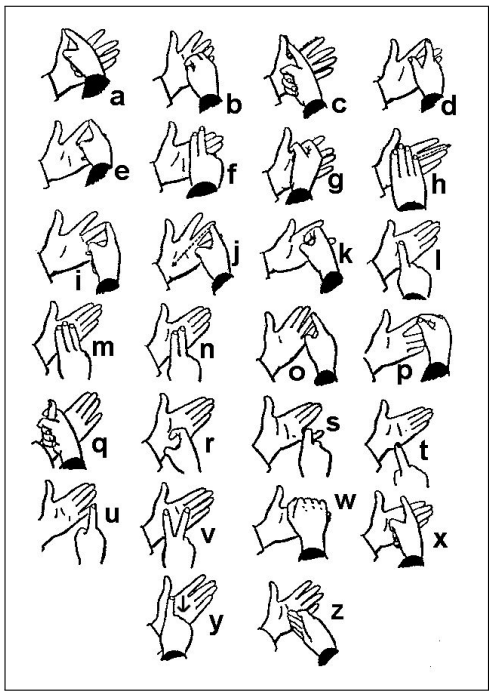
#### 1.3.2.4 Tactile Sign Languages (for DeafBlind people)

- **Definition**

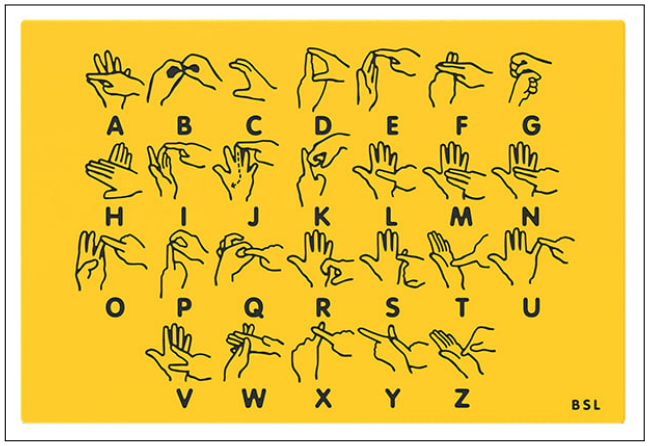
  According to Collins and Petronio (1998), in their chapter "What Happens in Tactile ASL?" published in the book Pinky Extension and Eye Gaze: Language Use in Deaf Communities, tactile sign languages are adaptations of visual sign languages in which DeafBlind individuals receive information through touch instead of sight.[14].

- **Examples**

1. **Tactile Sign Language (TASL,TBSL):** A touch-based adaptation of ASL or BSL for Deaf Blind individuals[12]. Signs are perceived by placing hands on the signer's hands or body. While retaining ASL or BSL grammar and vocabulary as shown in figure 1.1, visual elements like facial expressions are replaced by tactile cues such as hand pressure and movement, enabling effective communication in daily life and education[13].




(a) TASL




(b) TBSL

**Figure 1.1:Tactile Sign Languages: TASL (a) and TBSL (b)**

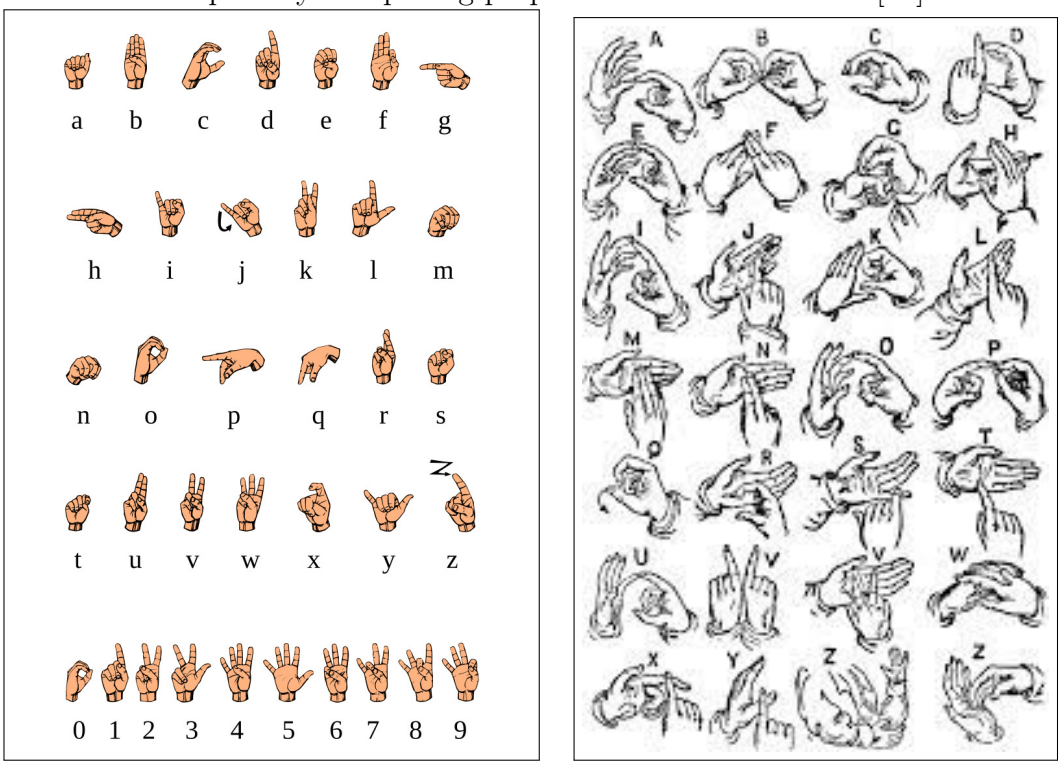
### 1.3.2.5 Finger-spelling/ Manual Alphabets

- **Definition**

  According to Daniels and Bright (1996) in their book "Sign Systems of the World", finger spelling is a system of hand shapes used to represent written letters of the alphabet and is therefore part of many sign languages but is not a language in itself[19].

- **Examples**

  1. **American Manual Alphabet (AMA):**it used in American Sign Language (ASL) to represent English letters through finger formations with one hand as shown in figure 1.2 (a). It is typically used to spell out names, places, technical terms, and words that do not have specific signs[7].
  2. **British Two-Handed Manual Alphabet (BSL Manual Alphabet):**This alphabet is used in British Sign Language (BSL), and it represents letters using both hands instead of one as shown in figure 1.2 (b). Many letters are formed through specific interactions between the dominant and non-dominant hands. It is used especially for spelling proper names and loanwords[13].




(a) ASL (b) BSL

**Figure 1.2:Manual Alphabets: ASL (a) and BSL (b)**

### 1.3.3 Uses of sign language

- **Primary Communication for Deaf and Hard of Hearing:** Enables effective communication for individuals who are Deaf or hard of hearing, facilitating daily conversation, education, and social interaction.
- **Education:** Used as a medium of instruction in schools for Deaf students, helping literacy, cognitive development, and learning of other languages.
- **Accessibility:** Ensures access to public services, TV programs, and online content. Sign language interpreters make medical, legal, and emergency information accessible.
- **Cultural and Social Identity:** Maintains and promotes Deaf culture and identity, acting as a social connector within the Deaf community.
- **Two-Way Communication:** Facilitates communication between Deaf and hearing people, bridging gaps in workplaces, schools, and social settings.
- **Emergency Situations:** Useful for non-verbal communication when voice is not possible, e.g., in noisy environments or during disasters.
- **Cognitive and Language Development:** Supports language acquisition in children who cannot hear and encourages spatial awareness, memory, and visual learning skills.
- **Inclusion and Advocacy:** Promotes inclusive societies and allows equal participation of Deaf people in social, educational, and professional activities; also used in advocacy campaigns.

### 1.3.4 Communication Problems in Sign Language

- **Limited Mutual Understanding:** Different countries and regions have their own sign languages (e.g., ASL, BSL, LSF), making cross-regional communication difficult.
- **Lack of Interpreters:** Professional sign language interpreters are often unavailable in public services, hospitals, or legal settings, creating communication barriers.
- **Low Awareness Among Hearing People:** Many hearing individuals do not know sign language, which limits interaction and inclusion for Deaf individuals.
- **Technological Barriers:** Not all technology (e.g., video calls, online education platforms) supports clear visual communication or interpretation.
- **Educational Inequality:** Some schools lack sign language instruction or resources, which can hinder the learning process for Deaf students.
- **Cultural Misunderstandings:** Hearing people may not fully understand the cultural aspects of Deaf communication, leading to misinterpretation or exclusion.
- **Incomplete Recognition:** In many countries, sign language is not legally recognized as an official language, reducing accessibility and rights for Deaf communities.

## 1.4 Existing Sign Language Databases

Numerous sign language databases are available for research purposes, particularly in fields such as computer vision and natural language processing. These databases vary depending on the sign language, the type of data collected (videos, images, key points), and their availability (open/restricted). We can categorize these databases into five types based on their intended use:

### 1.4.1 Isolated Sign Language Recognition Dataset (ISLRD)

Is a datasets designed to train models to classify a single sign (word) in isolation, with no temporal context between multiple signs. Such datasets are used in applications like sign language dictionaries, educational tools, and communication tools.

- **Examples**
  1. **ASL Citizen (USA):**This extensive, community-based collection contains 83,399 video samples of 2,731 distinct signs from American Sign Language. These were recorded by 52 different participants, 49 of whom identify as deaf and 3 as hard of hearing. Its key design feature is the use of realistic environments, resulting in natural variations in lighting, background, and user appearance, which are readily available for scientific research(Open Access)[15].
  2. **Alabib-65 (Algerian Sign Language, Algeria):**This restricted-access (permission-required) resource contains 6,238 video samples focusing on 65 distinct signs of Algerian Sign Language (SLA), performed by 41 sign language users. Primarily designed for Isolated Sign Language Recognition (ISLR) research, this resource provides high-resolution data collected in controlled environments to ensure quality[16].

### 1.4.2 Continuous Sign Language Recognition Dataset (CSLRD)

Recognizing a series of signs in a video clip is called CSLR. These signs can be linked together to form words, phrases, or complete sentences, creating a database of sign sequences that represent specific sentences, called CSLRD. It differs from an isolated sign language recognition database (ISLRD) in that it treats temporary joint expression, meaning that the movements of one sign affect the movements of the next sign.

- **Examples**
  1. **ArabSign:**is a valuable,restricted-access (permission-required) multimedia resource. With 9,335 sentence-level video samples from various participants, it provides deep maps, synchronized RGB video, and 3D joint data—a rich input stream necessary for fluent Arabic Sign Language (ArSL) temporal modeling[17].
  2. **How2Sign (ASL):**It is one of the largest and most comprehensive ongoing,(open-access) sign language databases developed for American Sign Language (ASL) translation and multimedia research. The database contains over 79 hours of sign language, captured from multiple camera angles to improve spatial and phonetic analysis[20].

3. **RWTH-PHOENIX-Weather (German Sign Language, Germany):**is a large, continuous,restricted-access (permission-required) dataset for German Sign Language (DGS), consisting of 6,841 sentence-level video sequences performed by nine individuals using a vocabulary of approximately 1,295 terms. The videos were recorded in a realistic weather forecast environment to capture natural expressions and variation in sign language users' performance (open access). [18].

### 1.4.3 Sign Language Translation Dataset (SLTD)

These are datasets used to convert continuous sign language videos directly into natural language text. They contain parallel sets of CSLRD data, where each sign language video is aligned at the sentence level with its corresponding text translation. These datasets provide the temporal and linguistic structure necessary for learning comprehensive translation models. However, because they contain fixed sentence pairs, this often results in SLT systems lacking flexibility .

- **Examples**
  1. **How2Sign (ASL):**It is one of the largest and most comprehensive ongoing,open-accesssign language databases developed for American Sign Language (ASL) translation and multimedia research. The database contains over 79 hours of sign language, captured from multiple camera angles to improve spatial and phonetic analysis[20].
  2. **CSL-Daily (Chinese Sign Language, China):** is a large-scale, continuous (open-access) dataset for Chinese Sign Language, containing 20,654 sentence-level video sequences recorded from ten signers with a vocabulary of approximately 2,000 glosses. The dataset includes gloss-level annotations and corresponding Chinese text translations, supporting both Continuous Sign Language Recognition and Sign Language Translation tasks [19].

### 1.4.4 Finger-spelling / Alphabet Recognition Dataset

It is a collection of videos or images that capture the hand gestures corresponding to individual letters in the sign language alphabet. Each gesture represents a single letter. Unlike ISLR datasets that focus on whole words, this dataset focuses on letter level recognition, which is necessary for representing proper nouns, technical terms, or words that do not have standard signs.

- **Examples**
  1. **ArASL(Arabic Sign Language Alphabet):**ArASL is an open-access character-level dataset, and recognized character sequences can be combined to form meaningful words or phrases. This allows models trained on ArASL to serve as building blocks for generating dynamic words and sentences.[21].

2. **ASL(American Sign Language Alphabet):**It is an open-access dataset containing approximately 87,000 RGB images representing 26 letters of the English alphabet, collected from more than 50 sign language speakers. Each letter is performed multiple times to capture variations in hand shape, orientation, and movement speed.[22].

### 1.4.5 Multi-Modal / 3D Sign Recognition Dataset

These are sets of recordings of sign language that incorporate various modalities, including 3D poses, RGB video, depth maps, and skeleton data. These datasets allow for more accurate recognition of complex gestures because, in contrast to standard 2D datasets, they capture spatial and temporal information about the signer's hands, body, and occasionally facial expressions.

- **Examples**
  1. **ChaLearn LAP Continuous Gesture Dataset:**It is an open-access dataset containing thousands of videos captured in several ways: RGB video, depth maps, 3D skeleton joints, and infrared. The dataset includes more than 13,000 gesture sequences performed by approximately 27 participants, covering discrete and continuous gestures[23].
  2. **mArSL:**This is a restricted-access (permission required) multimedia dataset for 3D Arabic sign language, containing separate signs performed by 4 participants aged between 21 and 30. The dataset contains 6,748 video samples covering 50 unique Arabic signs[24].

## 1.5 Existing and Domains of Assistive Applications

### 1.5.1 Hand Talk

- **Definition**

  It is a sign language translation system and AI-based educational tool that works online on mobile phones and uses machine learning models , such as 3D computer graphics (avatars), to display the complex and dynamic linguistic structures of different sign languages[25].




**Figure 1.3:Hand Talk Application**

- **Application Domains**

  This app can be used in different application domains, such a:

  1. **Education:**Used as a support tool in training settings and educational institutions to translate spoken or written content into sign language for Deaf students.
  2. **Public Accessibility:**Implemented in public institutions (hospitals, government offices, airports) to provide instant translation and improve communication with Deaf users.
  3. **Customer Service and Business:**Used by companies to facilitate communication with Deaf clients, especially in service centers, reception desks, and online platforms.
  4. **Health Services:**Helps medical staff communicate basic information to Deaf patients, improving clarity and accessibility in healthcare settings.

### 1.5.2 Lingvano

- **Definition**

  It is an online learning platform for sign language , designed for users of American Sign Language (ASL) and other sign languages through interactive video lessons. The platform focuses on accessibility and learning at a pace that suits each learner by offering short, practical lessons led by deaf instructors[26].




**Figure 1.4:Lingvano Application**

- **Application Domains**

  This app can be used in different application domains, such as:

  1. **Language Education:**Used as a structured digital platform for learning American Sign Language (ASL) and other sign languages through interactive, step-by-step lessons.
  2. **Community Accessibility:**Supports accessibility initiatives in schools, hospitals, and public institutions by enabling staff and users to acquire basic sign-language communication skills.
  3. **Health Services:**Help the Medical staff learn how to communicate essential information to deaf patients, improving clarity and accessibility in healthcare settings.
  4. **Family and Caregiver Support:**Helps family members and caregivers learn essential signs to communicate effectively with Deaf or Hard-of-Hearing relatives.

### 1.5.3 ASL Dictionary (Software Studios)

- **Definition**

  It is a comprehensive video-based dictionary that provides visual demonstrations of thousands of signs in American Sign Language , including words, phrases, and idiomatic expressions[27].




**Figure 1.5:ASL Dictionary Application**

- **Application Domains**

  This app can be used in different application domains, such as:

  1. **Language Education:**Used by students, educators, and self-learners to study and practice ASL vocabulary and grammar.
  2. **Reference Tool:**Provides quick and accurate sign language search for interpreters, teachers, and learners in everyday or educational communication contexts.
  3. **Communication Support:**Assists hearing individuals and deaf users in real-world conversations by providing a reliable resource for confirming or learning signs.

### 1.5.4 Sorenson ntouch (VRS)

- **Definition**

  It is an application for mobile phones and desktop computers that allows users to connect to Sorenson's Video Relay Service, enabling deaf and hard-of-hearing people to make video calls interpreted in American Sign Language (ASL) in real time . Through the app, users connect with professional American Sign Language (ASL) interpreters who translate between sign language and spoken English (or Spanish)[28].




**Figure 1.6:Sorenson ntouch Application**

- **Application Domains**

  This app can be used in different application domains, such as:

  1. **Communication:**Enables conversational flexibility: calls can be answered or transferred across devices (mobile, desktop).
  2. **Work and Professional Life:**Used in workplace settings (meetings, client calls) to provide full accessibility for Deaf employees.
  3. **Education:**Supports students who are Deaf or Hard-of-Hearing by enabling video calls with interpreters from school or home.
  4. **Accessibility in Public Services:**Used in government or public service settings to provide inclusive communication for Deaf users (e.g., in public offices).

## 1.6 Necessity of Implementing Recent Applications

Modern assistive applications that facilitate communication, education, and social participation for deaf and hard-of-hearing people have been made possible by recent technological advances. Applications such as **Hand Talk**, **Lingvano**,**ASL Dictionary**, and **Sorenson ntouch** use interactive interfaces, video recognition, and artificial intelligence to improve accessibility in a variety of areas, including everyday communication, healthcare, and education. Despite these technological advances, the use of these apps remains limited, underscoring the need to consider the benefits, challenges, and necessity of implementing these modern tools [29]:

1. **Advancing Communication Equity:**The efficacy of contemporary assistive sign-language applications in mitigating long-standing communication disparities represents a compelling rationale for their development and deployment. In their systematic review, Correa assert that inclusive technologies incorporating sign language facilitate successful bilateral communication between the Deaf and hearing communities, thereby validating sign language as a legitimate mode of communication rather than an impediment.
2. **Social Inclusion and Cultural Respect:** The implementation of modern applications is not limited to ease of use, but also includes the recognition of sign language as a full language and the identity of deaf people. That digital tools can help transform language imposition into respect for deaf culture, promoting not only communication but also cultural integration.
3. **Critical Need in Healthcare and Public Services:**the outdated nature of communication infrastructure calls for the mandatory application of new technology solutions to ensure accessible and equal Sign Language services. This need is imperative in overcoming present communication barriers and ensuring linguistic inclusion. Notably, AI solutions, including recognition and translation, are vital in addressing Audistic norms, particularly in industries such as Healthcare and Public Services. Finally, technological upgradation is imperative in ensuring linguistic human rights and Deaf community inclusion in Cyberspace.

## 1.6 Conclusion

In this chapter, we describe the key notions related to sign languages such as definition, classification, typology, and application areas. Moreover, we stress the major problems connected with the communication issues in this field, particularly those related to the interaction of sign language speakers with people who use spoken languages. The chapter also describes some existing datasets of sign languages and discusses their importance for practical applications development. The chapter provides the necessary background information for understanding the scope of this study.The next chapter focuses on learning techniques, aiming to identify the methods used and explain their mechanisms of action. It also presents a review of related work in this field.

# Chapter2- Learning Techniques

## 2.1 Introduction

These kinds of technologies are trying to reproduce some aspects of human thought processes in order to develop computers capable of learning, improving their decision-making ability, solving hard problems or performing tasks that usually need human-level reasoning[30].

The current chapter will be dedicated to using artificial intelligence in sign languages. First, the chapter discusses the most basic learning approaches. Learning approaches may involve using model-based algorithms, such as traditional machine learning and deep learning that are built around training of models with annotated data that allow recognizing patterns, classifying gestures, and predicting sequences. This is the basis of all automated sign language recognition algorithms. Moreover, AI paradigm-based learning approaches will emphasize more recent techniques that involve the use of large language models (LLMs), as well as other computer vision algorithms that are based on pre-training, self-supervision, and enhanced reasoning skills. Following this theoretical discussion, the chapter will describe the examples of using AI in sign language applications, listing sources used, datasets, and results obtained.

## 2.2 Model-based Training Techniques

In Sign Language Recognition, the choice of learning technique plays a critical role in determining system accuracy and robustness. Different model-based approaches have been developed to capture both static and dynamic gestures, each with its own strengths and limitations.

### 2.2.1 Machine Learning Techniques

#### 2.2.1.1 Support Vector Machines (SVM)

- **Definition**

  Support Vector Machines (SVM) are a class of supervised machine learning models used for classification and regression tasks. As described in Artificial Intelligence: A Modern Approach by Russell and Norvig, SVMs operate by identifying an optimal separating boundary (hyperplane) in the feature space that maximizes the margin between different classes. This enables SVMs to provide strong generalization performance by focusing on the most informative data points, known as support vectors, which define the decision boundary[30].

- **Mechanism of Action**

  The SVM approach represents each input object as a point in a high-dimensional feature space. Then, it finds an optimum hyperplane in the space separating the classes with maximum margin between the classes, which refers to the distance from the hyperplane to the closest points of both classes (called support vectors). This is demonstrated in figure 2.1.

When dealing with non-linearly separable datasets, the support vector machines use certain kernel functions like polynomial or RBF (Radial Basis Function) to project the dataset to a high dimensional space where the problem becomes linearly separable. Since SVM is a binary classifier by nature, a multiclass classification problem (for instance, recognizing several types of hand gestures) requires some method for combining multiple binary SVM classifiers. This may be achieved via techniques such as One-vs-Rest or One-vs-One, whereby a separate SVM model is built for each class versus other classes (the former case) or every pair of classes (the latter)[32].

In prediction, each new input sample is assigned to the class corresponding to the side of the separating hyperplane to which it belongs.[33].

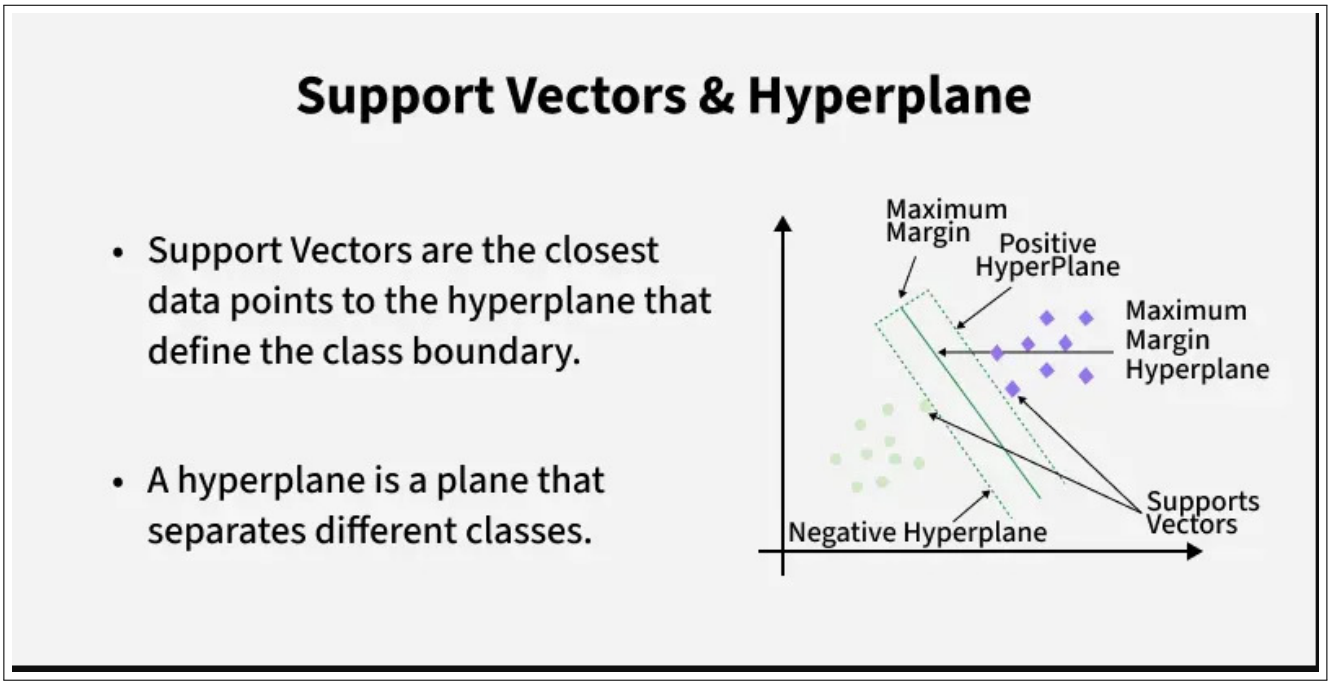



**Figure 2.1: SVM Architecture**

### 2.2.1.2 Multi-Layer Perceptron (MLP)

- **Definition**

  A Multi-Layer Perceptron (MLP) is a type of feedforward artificial neural network consisting of multiple layers of interconnected neurons with nonlinear activation functions. It is trained using backpropagation to adjust its weights so as to model complex, nonlinearly separable data patterns[34].

- **Mechanism of Action**

  In a neural network, the input data is first represented as a feature vector, which may include descriptors such as hand shape features, keypoints, or other extracted visual features, and is provided to the input layer of the network.

Each neuron in the hidden layers computes a weighted sum of its inputs and adds a bias term, forming the neuron's net activation . This activation is then passed through a nonlinear activation function, such as ReLU, sigmoid, or tanh, which enables the network to capture and model complex nonlinear relationships within the data . The outputs of one layer are propagated forward to the subsequent layer in a process known as forward propagation, continuing through all hidden layers until the output layer is reached[34] .

The output layer generates the final prediction, and for classification tasks, functions such as softmax or sigmoid are typically applied to produce class probabilities . The predicted output is compared with the true target labels using a loss function, for instance, cross-entropy, which quantifies the classification error . This error is then propagated backward through the network using the backpropagation algorithm to compute gradients of the loss with respect to each weight [35]. Network weights and biases are updated iteratively using gradient-based optimization methods, such as gradient descent, to minimize the loss as shown in figure 2.2. Steps of weighted summation, activation, forward propagation, error computation, backpropagation, and weight update are repeated over multiple training epochs until the network converges and achieves satisfactory classification performance.

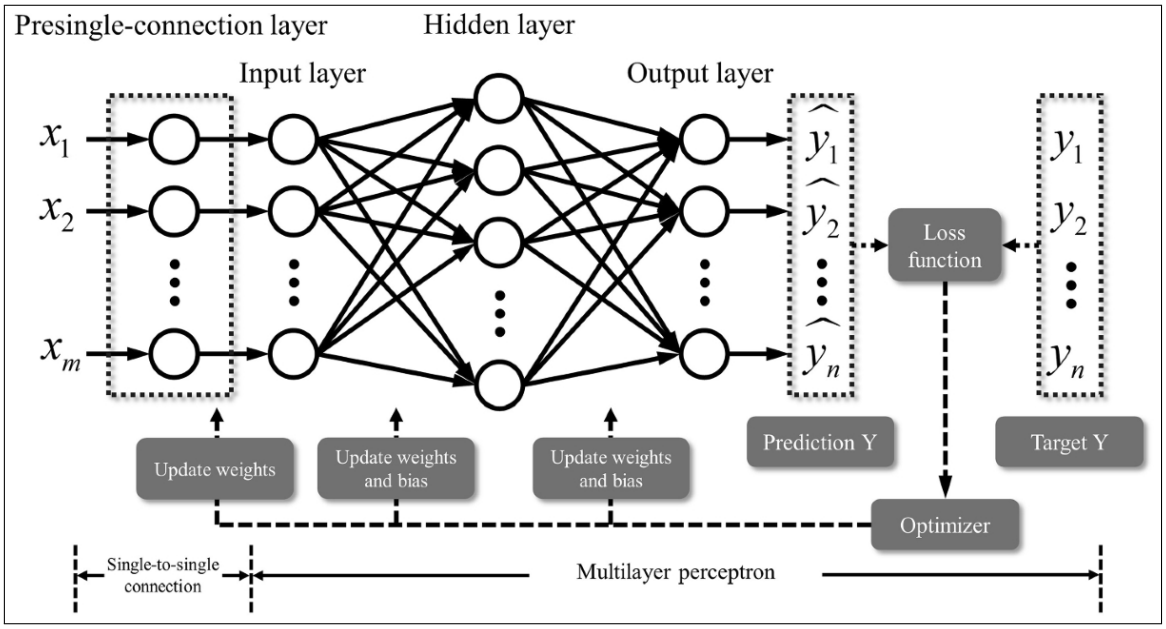



**Figure 2.2:MLP Architecture**

### 2.2.1.3 Decision Trees / Random Forests

- **Definition**

  Decision Trees are a type of supervised learning model that recursively partitions the input space into regions corresponding to different output classes. Each internal node of the tree tests a feature, each branch represents an outcome of the test, and each leaf node assigns a class label. The tree is constructed by selecting features that maximize information gain or reduce impurity, enabling a hierarchical, interpretable classification of instances[36].

Random Forests extend decision trees by constructing an ensemble of trees using random subsets of features and training samples, then aggregating their predictions (typically via majority voting). This approach improves generalization and robustness compared to a single decision tree, reducing overfitting while maintaining high classification accuracy.

- **Mechanism of Action**

The methods of decision tree and random forest use the input dataset that is represented as a feature vector, containing both continuous and categorical variables. In the case of decision trees, the recursive partitioning process takes place in each node depending on certain characteristics like Gini impurity, entropy, or variance reduction, resulting in the development of branches and the determination of outputs at leaf nodes using either the most frequent class or average values [37]. The process of random forests uses bootstrapping and the creation of different trees that are trained with randomly chosen subsets of features, which improves the diversity and decreases correlations between individual trees. The output of random forest models is determined through aggregation of the outputs of individual decision trees using majority voting and averaging for classification and regression tasks, respectively,as shown in figure 2.3.

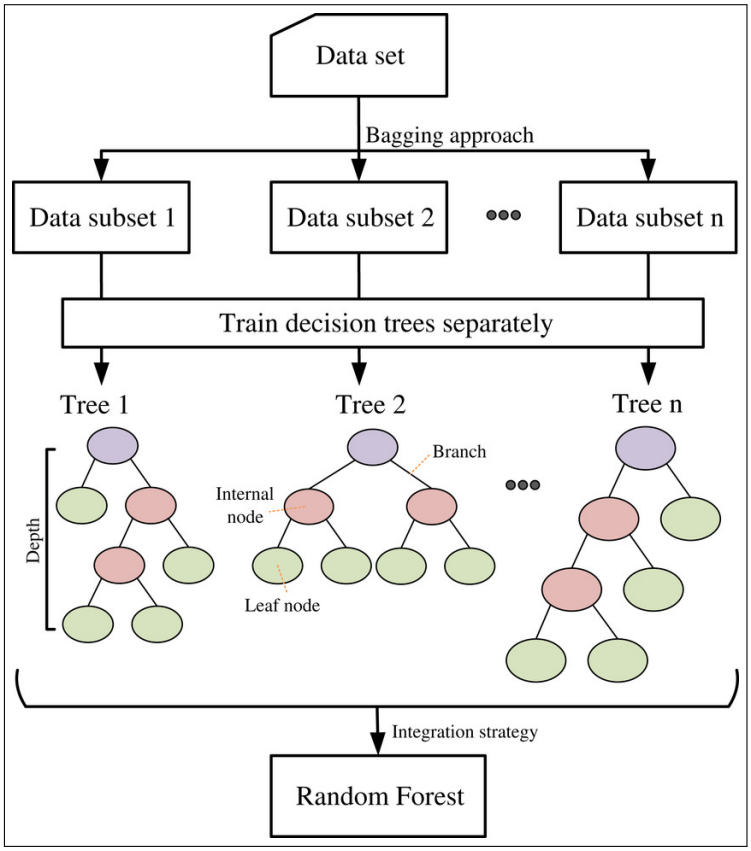



**Figure 2.3:DT and RF Architecture**

## 2.2.2 Deep Learning Techniques

### 2.2.2.1 Convolutional Neural Network(CNN)

- **Definition**

  Convolutional Neural Networks (CNNs) is a specialized class of deep neural networks designed to process data with a grid-like structure, such as images. CNNs exploit local spatial connectivity and parameter sharing through convolutional filters, allowing them to automatically learn hierarchical feature representations directly from raw input data. By combining convolutional layers, non-linear activation functions, and pooling operations, CNNs are able to capture spatial patterns while reducing sensitivity to translation and distortion[38].

- **Mechanism of Action**

  A convolutional neural network (CNN) processes an input image or video frame as a network of pixel values. Convolutional filters extract low-level spatial features, followed by nonlinear activations such as ReLU to enhance representational power. Pooling layers reduce spatial dimensions while preserving essential information, providing translation stability[30]. Stacked convolutional and pooling layers gradually learn high-level features, such as hand shapes and spatial configurations, to recognize gestures. These features are then flattened and passed to fully connected layers for classification. The output layer produces class probabilities, and the network is trained using gradient descent and optimization to minimize classification loss, enabling accurate recognition of complex visual patterns as shown in figure 2.4 .

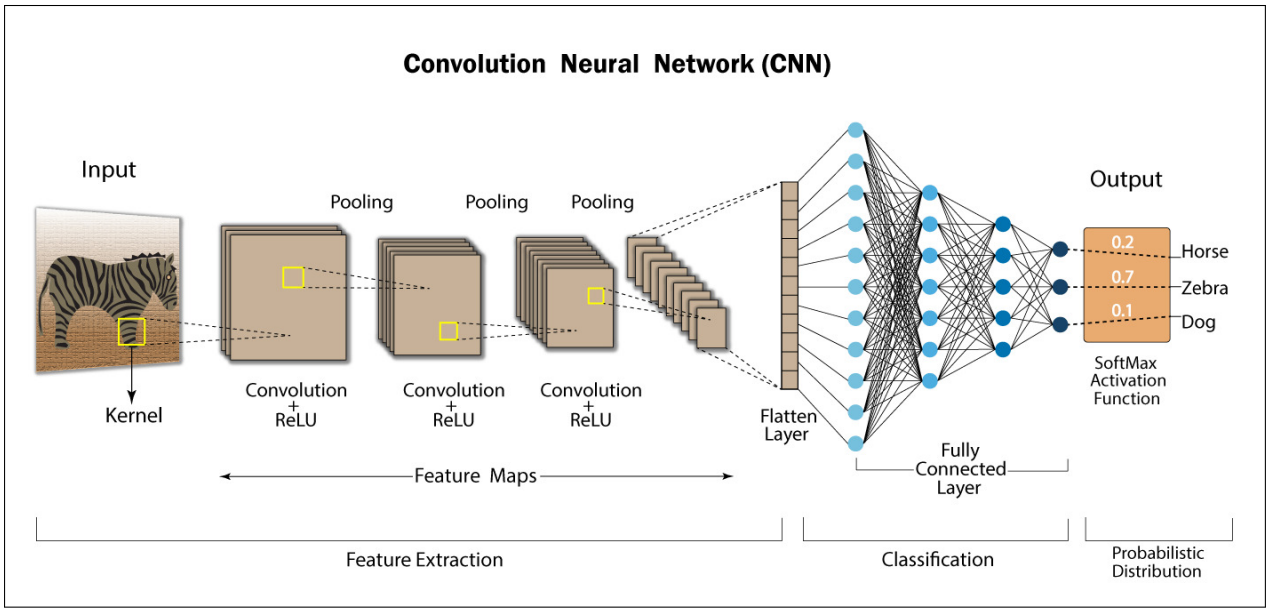



**Figure 2.4:CNN Architecture**

### 2.2.2.2 Recurrent Neural Networks (RNNs)

- **Definition**

  Recurrent Neural Networks (RNNs) are a class of neural network models designed to process sequential data by maintaining an internal state that captures information from previous inputs. RNNs differ from feedforward neural networks in that they incorporate feedback connections, allowing the network to use past information to influence current outputs. This structure makes RNNs particularly well suited for tasks involving temporal dependencies, such as speech recognition, natural language processing, and time-series prediction[39].

- **Mechanism of Action**

  Recurrent Neural Networks (RNNs) process input data sequentially, where each element corresponds to a specific time step. At each step, the network updates its hidden state by combining the current input with the previous hidden state through weighted connections[39]. This hidden state serves as a memory, preserving information from past inputs and enabling the modeling of temporal dependencies. The network generates outputs either at each time step or after processing the entire sequence based on the updated hidden state. During training, errors are propagated backward through time using Backpropagation Through Time (BPTT), and network parameters are optimized with gradient-based learning methods to minimize the loss function as shown in figure 2.5.

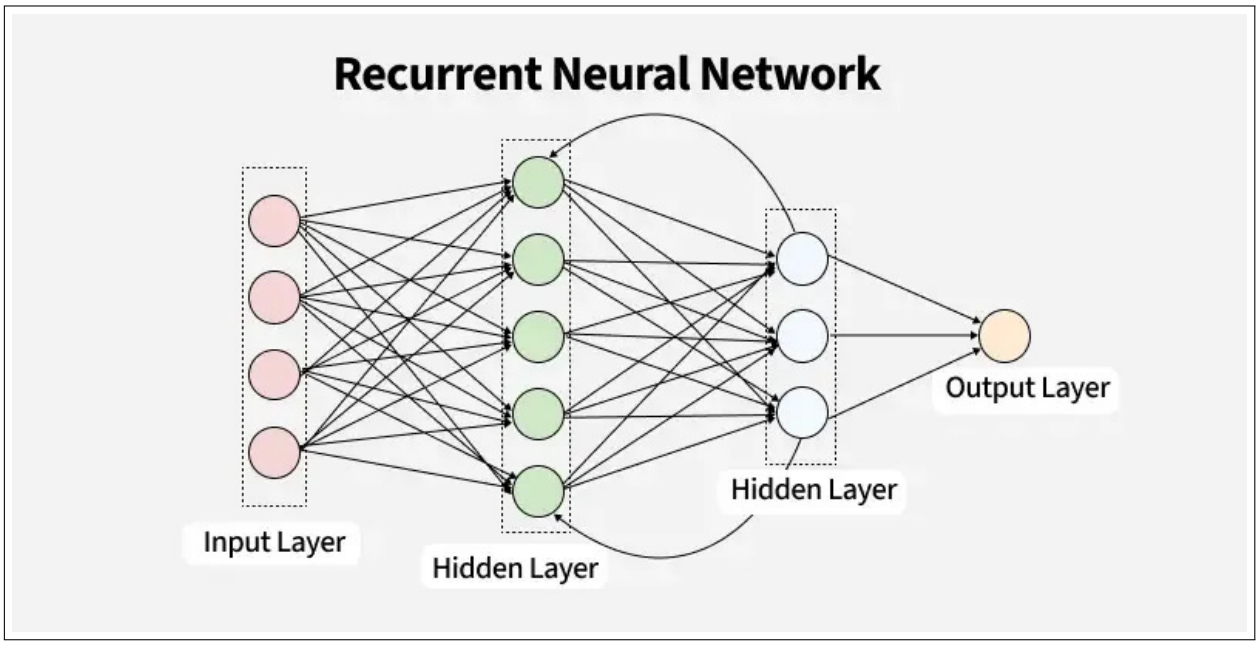



**Figure 2.5:RNN Architecture**

### 2.2.2.3 Long Short-Term Memory (LSTM)

- **Definition**

  Long Short-Term Memory (LSTM) is a specialized type of recurrent neural network designed to address the vanishing and exploding gradient problems encountered in standard RNNs when learning long term dependencies. LSTM networks introduce a memory cell and a set of gating mechanisms namely the input gate, forget gate, and output gate that regulate the flow of information, allowing the network to selectively retain or discard information over extended sequences. This architecture enables LSTMs to effectively model temporal patterns in sequential data such as speech, handwriting, and gesture sequences[40].

- **Mechanism of Action**

  The operation of a Long Short-Term Memory (LSTM) network processes sequential input data over time while maintaining an internal memory state. At each time step, the forget gate selectively removes irrelevant information from the previous cell state, while the input gate determines which new information should be incorporated. The cell state is then updated to preserve long-term contextual information. Subsequently, the output gate controls the exposure of the cell state to produce the current hidden state, which is propagated to the next time step or subsequent layers. During training, learning is performed through backpropagation through time[40], enabling the network to optimize its parameters and effectively model long-range temporal dependencies in sequential data as shown in figure 2.6.

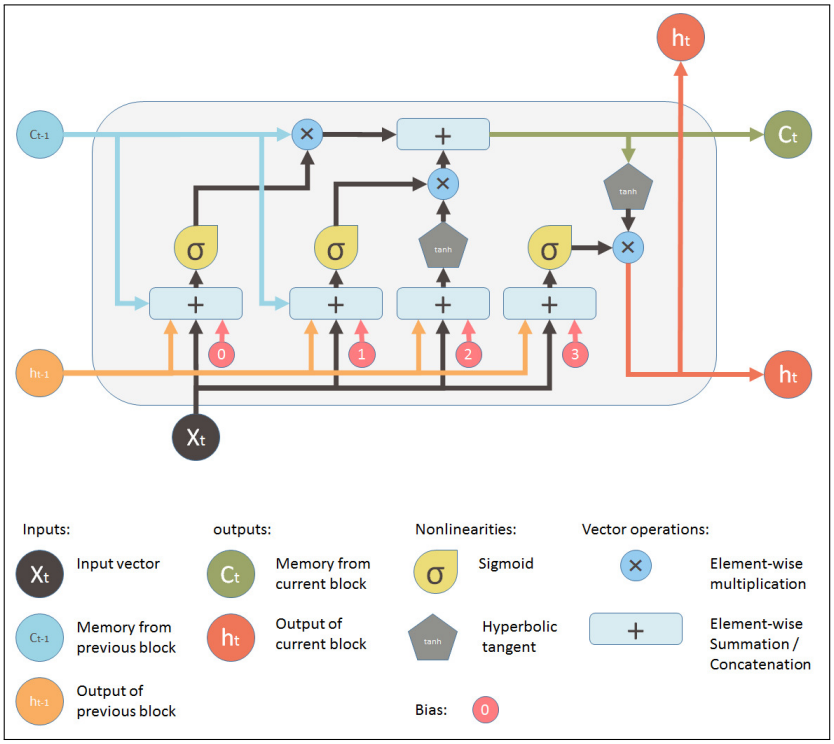



**Figure 2.6:LSTM Architecture**

## 2.3 AI Paradigm-based Techniques

### 2.3.1 Large Language Models (LLMs)

- **Definition**

  Large Language Models (LLMs) are advanced neural network–based models designed to learn and represent the statistical structure of natural language from large-scale textual corpora. such models operate within the learning agent paradigm and rely on probabilistic reasoning to handle uncertainty inherent in real-world domains such as human language.

  LLMs are typically built using transformer-based architectures that employ self-attention mechanisms to capture long-range dependencies and contextual relationships between tokens in a sequence. This design enables LLMs to perform a wide range of natural language processing tasks, including text generation, question answering, translation, and summarization, with a high degree of contextual coherence and adaptability[41].

- **Mechanism of Action**

  In the Transformer architecture, the input text is initially converted into a sequence of discrete tokens using a predefined tokenization scheme. These tokens may correspond to words, subwords, or characters, allowing the model to process the text in a structured numerical form [41]. Each token is then mapped to a continuous vector representation known as a token embedding. Because the Transformer does not inherently capture sequence order, positional encodings are added to the embeddings, ensuring that the model retains information about the relative positions of tokens within the sequence.

  The embedded tokens are processed by a multi-head self-attention mechanism, which computes relationships between tokens and assigns attention weights to capture long-range dependencies and contextual information. Each attention head focuses on different linguistic aspects, and their outputs are combined to form a comprehensive representation of the sequence.

  Following attention processing, the representations are passed through a position wise feed forward neural network, which applies non linear transformations to further increase the expressiveness and modeling capacity of the network. To stabilize training and improve information flow, each sublayer both the attention and feed-forward layers—is augmented with residual connections and layer normalization. Multiple such layers are stacked to progressively refine the token representations, enabling the model to encode increasingly sophisticated contextual patterns.

  Finally, the Transformer produces outputs through a probabilistic prediction layer, which estimates the likelihood of the next token in the sequence. By sampling or selecting the token with the highest probability, the model generates coherent and contextually appropriate sequences of text.

This architecture allows the Transformer to capture both local and global dependencies efficiently, making it highly effective for a wide range of natural language processing tasks, including translation, summarization, and sequence prediction As shown in figure 2.7.

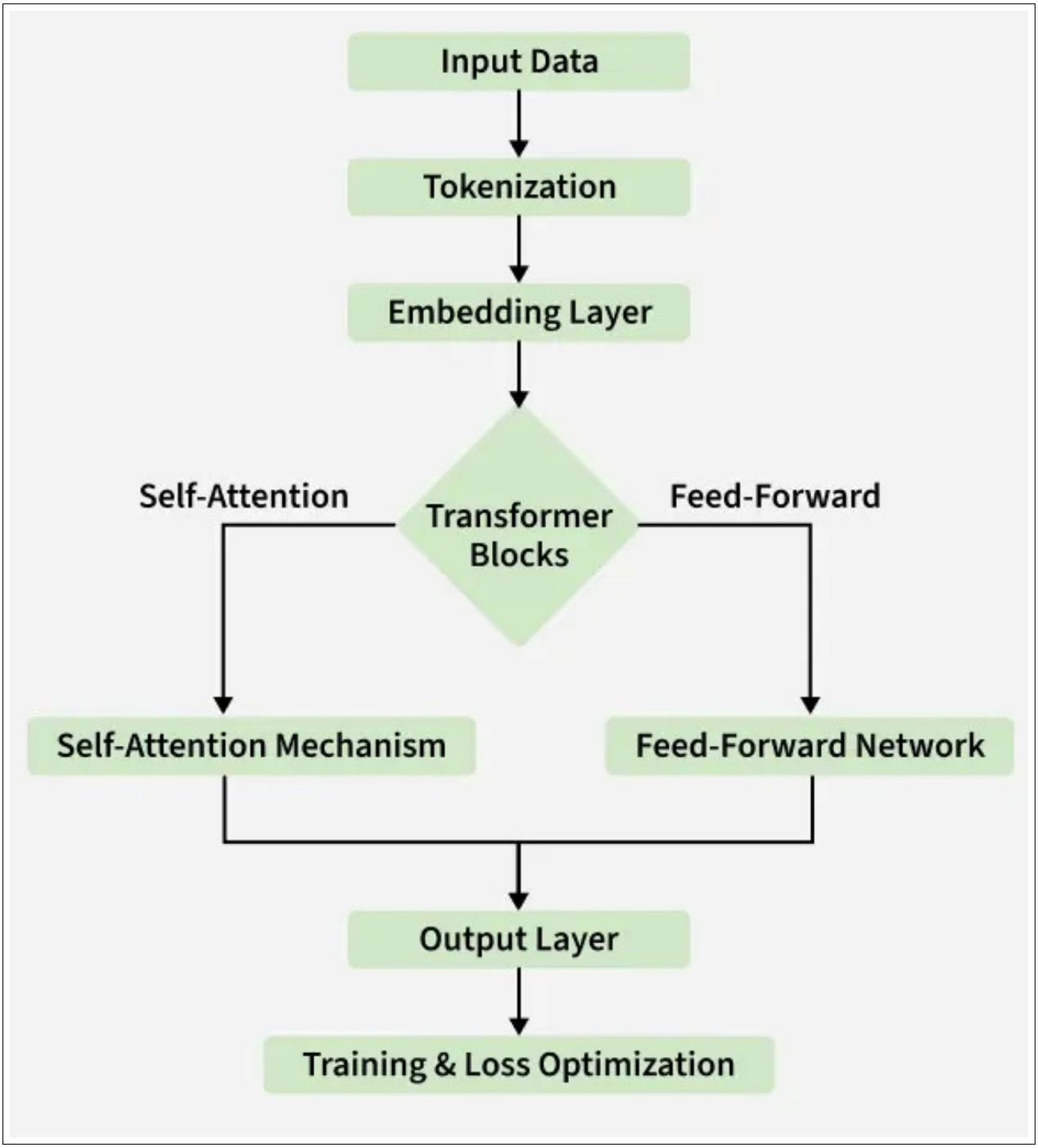



**Figure 2.7:LLM Architecture**

- **Overview of Contemporary LLMs**
  1. **GPT (Generative Pre-trained Transformer) Series – OpenAI:**Includes GPT-2, GPT-3, and GPT-4. These are autoregressive models trained to predict the next token in a sequence and are widely used for text generation, summarization, and conversational applications[42].
  2. **Gemini – Google DeepMind:**A recently developed LLM by Google DeepMind that extends traditional transformer-based models with enhanced reasoning abilities, multimodal understanding (e.g., text and image integration), and advanced safety alignment mechanisms to improve robustness and reduce undesirable outputs in complex language tasks[43].
  3. **BERT (Bidirectional Encoder Representations from Transformers) – Google:** An encoder-only model that understands context bidirectionally. Commonly applied in question answering, text classification, and sentiment analysis[44].
  4. **T5 (Text-to-Text Transfer Transformer) – Google:**Treats all NLP tasks as text-to-text problems, enabling applications such as translation, summarization, and classification[45].
  5. **PaLM (Pathways Language Model) – Google:**A large-scale model designed for reasoning, multitask capabilities, and code generation[46].
  6. **LLaMA (Large Language Model Meta AI) – Meta:**A lightweight, research-oriented model optimized for academic and experimental purposes[47].
  7. **Claude – Anthropic:**A safety-focused LLM designed with alignment and ethical AI considerations in mind[48].
  8. **Mistral, Falcon, and BLOOM:**Open-source LLMs developed for multilingual support and high-performance NLP tasks[49].

### 2.3.2 Transformer

- **Definition**

  The Transformer is a neural network architecture designed to model sequential data using self-attention mechanisms instead of recurrence, enabling efficient learning of long-range dependencies and contextual relationships. this architecture functions as a learning agent that processes input sequences, represents knowledge probabilistically, and generates outputs aimed at optimizing a defined objective. By leveraging multi-head attention, positional encoding, and stacked feed-forward layers, the Transformer provides a scalable and flexible framework for natural language understanding and generation[50].

- **Mechanism of Action**

  The transformer architecture serves as the heart of Large Language Models, providing the fundamental mechanisms for capturing long-range dependencies, contextual relationships[50], which are then scaled and leveraged by LLMs to perform a wide range of natural language processing tasks as shown in figure 2.8.

  We can divide the architecture into 6 basic parts where[51]:

Firstly, each token is mapped to a dense vector representation through the embedding layer. the embedding of tokens contains information about the basic meaning of the token.this way, the task is converted from working directly on tokens to working with their numerical representation.

Secondly, after adding positional encodings to the embedding, it becomes possible for the model to understand the order of the tokens inside the sequence.this is important because unlike recurrent models, transformers do not have built-in support for sequential ordering.

Thirdly, each embedded token passes through a linear transformation that produces three vectors – Query (Q), Key (K), and Value (V). These vectors play a crucial role in the attention mechanism:

$$\boxed{\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V}$$

Scaled dot-product attention is performed on Query, Key, and Value vectors through computing a weighted sum. The scaling factor and softmax layer normalize the results of the dot-product, making it possible to interpret these values as attention scores. Scores denote the importance of tokens regarding others, thus the values are multiplied by the value vectors.

Fourthly, scaled dot-product attention between tokens enables the attention between a particular token and the rest of the tokens in the sequence. In order to increase the capability of capturing a broader range of relations, several attention mechanisms are conducted concurrently, which are referred to as multi-head attention.

Fifthly, in order to stabilize learning and allow the gradients flow more freely, a residual connection of input tokens with attention scores output is used before normalization. after that, the result is passed through a position-wise feedforward neural network that uses two linear layers with a non-linear activation function to independently transform each token.

Sixthly, another residual connection and normalization step are performed to ensure stability and retain the information before passing it to subsequent attention blocks. this whole process is repeated many times to build several deep transformer layers.

Finally, the output of the last transformer block contains the context-aware representations of input tokens that are ready to be used in various downstream tasks like language modeling.

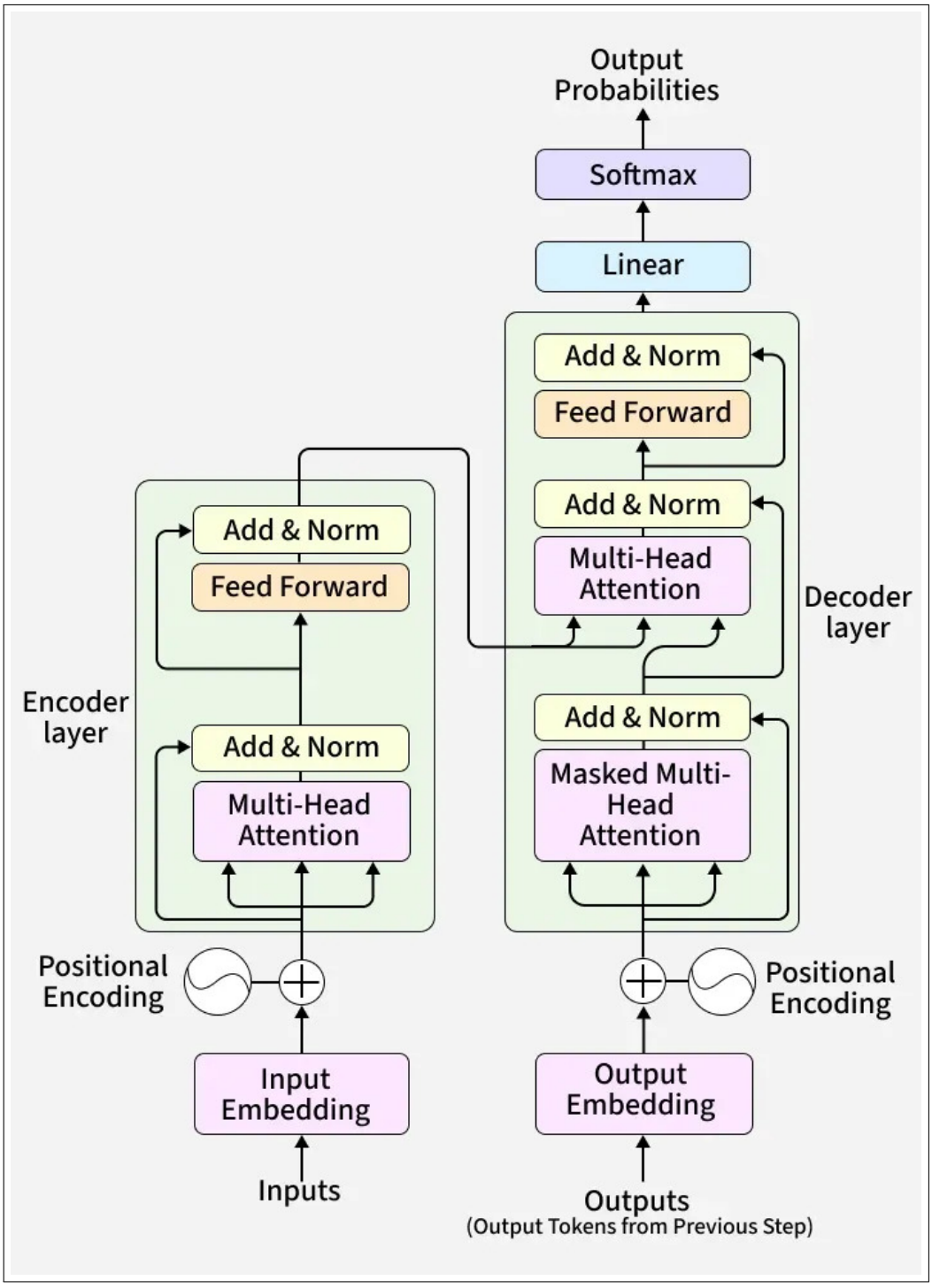



**Figure 2.8:Transformer Architecture**

### 2.3.3 Computer Vision (CV)

- **Definition**

  Computer Vision (CV) is a branch of Artificial Intelligence (AI) that helps computers to interpret and understand visual information much like humans. This tutorial is designed for both beginners and experienced professionals and covers key concepts such as Image Processing, Feature Extraction, Object Detection, Image Segmentation and other core techniques in CV[52].

- **Mechanism of Action**

  A typical computer vision system operates through a sequential pipeline that converts raw visual data into meaningful information as shown in figure 2.9. The main steps are as follows:

  Computer vision systems operate through a sequential pipeline that transforms raw visual input into meaningful information. The process begins with **Image Acquisition**, where visual data is captured from cameras or sensors. Next, **Preprocessing** improves image quality using noise reduction, normalization, and resizing. **Segmentation** then separates the image into meaningful regions or objects, followed by **Feature Extraction**, which identifies edges, corners, textures, and other important patterns. **Object Detection** locates and identifies objects within the scene, and **Object Recognition and Classification** assigns each object to a predefined category. Finally, **Post-processing and Visualization** refines the results and presents them for interpretation, decision-making, or interaction with other systems[53].

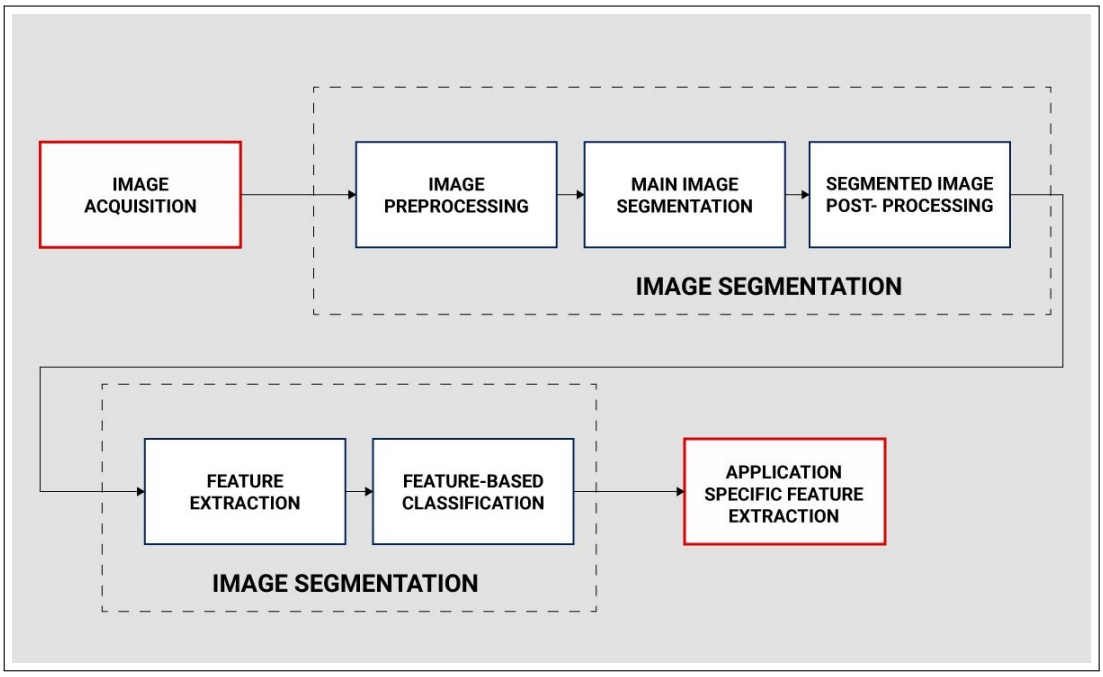



**Figure 2.9: Computer Vision Architecture**

- **Essential Tools and Libraries in Computer Vision**
  1. **OpenCV:** An open-source computer vision library providing a wide range of tools for image and video processing, including filtering, edge detection, object recognition, and camera calibration[54].
  2. **MediaPipe:** Developed by Google, MediaPipe enables building multimodal machine learning pipelines, offering ready-to-use solutions for hand tracking, face detection, and pose estimation[55].
  3. **PyTorch Vision:** Provides datasets, pre-trained models, and utilities for implementing deep learning–based computer vision tasks efficiently[56].
  4. **TensorFlow Lite:** A lightweight framework for deploying machine learning models on mobile and embedded devices, supporting efficient on-device computer vision applications[57].
  5. **OpenPose:** A library for real-time multi-person keypoint detection, including body, face, hands, and foot landmarks, widely used in pose estimation, gesture recognition, and human motion analysis[58].
  6. **Detectron2:** Developed by Facebook AI Research (FAIR), Detectron2 is a modular, high-performance library for object detection, instance segmentation, and keypoint detection. It is built on PyTorch and supports state-of-the-art models such as Mask R-CNN and Faster R-CNN[59].
  7. **OpenVINO:** Developed by Intel, OpenVINO (Open Visual Inference and Neural Network Optimization) is a toolkit for optimizing and deploying deep learning models for computer vision tasks on Intel hardware, including CPUs, VPUs, and FPGAs, supporting real-time applications[60].

## 2.4 Related Works

in this study, it is important to review related work to position our contribution within existing research. This section presents key approaches in gesture recognition and computer vision, ranging from traditional feature-based methods to modern deep learning techniques such as CNNs, RNNs,LSTMs, and hybrid models.
The following table 2.1 summarizes the main contributions, including methods, datasets, and performance results. This comparison helps identify the limitations of existing approaches and justifies the proposed method in this work.

| **Reference** | **Paper Title** | **Model / Technique** | **Dataset (Size, Classes)** | **Performance** | **Real-time?** |
|---|---|---|---|---|---|
| [61] | Real-Time Arabic Sign Language Translator Using MediaPipe and LSTM | BiLSTM + MediaPipe | Arabic Sign Language; 32 letters + 20–50 words + 10 digits (under 100 classes) | Letters 98%, Words 96%, Digits 100% | Yes |
| [62] | Isolated Video-Based Sign Language Recognition Using a Hybrid CNN-LSTM with Attention | MobileNetV2 + LSTM + Attention | WLASL dataset; 100 classes | Accuracy 84.65% | Not specified |

| | | | | | |
|---|---|---|---|---|---|
| [63] | Indonesian Language Sign Detection using Mediapipe with LSTM (ISLS)[62] | LSTM + MediaPipe Hands | 36 gestures (alphabetic and numeric); 25 sequences per gesture; Split: 80% train, 10% val, 10% test | Accuracy 97.1%, Precision 97%, Recall 96.6%, F1-score 96.4% | Yes |
| [64] | Real-Time Norwegian Sign Language Recognition Using MediaPipe and LSTM[63] | LSTM + MediaPipe | 1059 videos; 11 classes (numbers 0-10); 5 participants; 20 key frames/video | Testing accuracy 95% | Yes |
| [65] | Empowering Communication: A Deep Learning Framework for Arabic Sign Language Recognition with Attention [64] | LSTM + Attention + MediaPipe | 50 words and numbers; data from 5 volunteers | Accuracy ¿85% (individual), 83% (combined) | Not specified |
| [66] | Real-Time Arabic Sign Language Recognition Using a Hybrid Deep Learning Model [65] | CNN + LSTM hybrid | 20 words; 4000 images (10 static words) + 500 videos (10 dynamic words) | CNN accuracy 94.4%, LSTM accuracy 82.7% | Yes |
| [67] | Sign Language Recognition using Deep CNN with Normalised Keyframe Extraction and LSTM[66] | Deep CNN + LSTM | 10,000 videos; 50 sign classes | CNN accuracy 89.24%, LSTM accuracy 89.99% | Not specified |

Table 2.1 :Summary of Sign Language Detection Papers Using LSTM

## 2.5 Conclusion

This chapter has considered various learning techniques and approaches that are at the core of modern artificial intelligence technologies, giving special attention to machine learning and deep learning. This chapter goes on to analyze state-of-the-art approaches, such as language models, transformers, and computer vision techniques and explains their operations and uses. The survey of previous research has also been completed to provide the necessary background and to place this paper into an overall context. All together, the chapter forms a solid theoretical base for the creation of the suggested system.The next chapter, Application Design, focuses on the data description and preprocessing stage, followed by the presentation of the general architecture of the application. It also details the system design using UML diagrams, including use case, class, sequence, and activity diagrams, as well as the database design.

# Chapter3- Application Design

## 3.1 Introduction

Sign language recognition systems aim to overcome the challenges associated with communication between the deaf and hearing impaired by translating the input gestures into text or speech. Although considerable progress has been achieved in computer vision and deep learning, designing an application to recognize sign language remains an issue, given its intricate characteristics.

Sign language recognition faces several challenges, including its intricate spatiotemporal characteristics, which include motion, body posture, and other time-dependent factors. The system's performance is also affected by differences in users, speed, and environmental factors, including lighting and background. The availability of datasets is another factor that affects the design of sign language recognition systems, given their limited number of classes and diversity. Lastly, the output obtained from the system may not form grammatically correct sentences, which may require further processing to make it more coherent and readable.

To overcome these challenges, this chapter discusses the design of the proposed application, which includes an overview of the dataset used, justification of the models adopted, an overview of the design, and an in-depth discussion of the main components of the design. The discussion is intended to give readers an overview of how the application can be used to recognize sign language.

## 3.2 Data Description

The dataset used for the purpose of carrying out this study is the ASL Citizen Dataset. This dataset is a huge collection of data created for the purpose of carrying out American Sign Language (ASL) recognition. The dataset contains 83,399 video samples of 2,731 unique ASL signs made by 52 sign language users. On an average, 30 video samples exist for each sign language[15].

In the dataset, each sample is a short video clip in which a sign language user performs a unique American Sign Language sign. Each video sample contains spatial information, such as hand shapes, locations, and orientations, as well as temporal information, such as motion and dynamic gestures. This is essential for developing deep learning models for sign language recognition systems that rely on sequences or recurrent deep learning models[15].

A notable feature of the dataset is that it was collected using normal cameras in normal environments rather than using cameras in a controlled environment. As a result, the dataset is representative of a variety of lighting conditions, diverse scenes or backgrounds of sign language users, and different camera or user angles. This is an advantage compared to WLASL or MS-ASL since their videos were collected from the internet or other secondary sources(figure3.1)[15].

The dataset can be accessed at: ASL Citizen Dataset.

| Dataset | Vocab size | Videos | Videos/gloss | Signers | Collection | Consent |
|---|---|---|---|---|---|---|
| RWTH BOSTON-50 | 50 | 483 | 9.7 | 3 Deaf | Lab | ✓ |
| Purdue RVL-SLL | 39 | 546 | 14.0 | 14 Deaf | Lab | ✓ |
| Boston ASLLVD | 2,742 | 9,794 | 3.6 | 6 Deaf | Lab | ✓ |
| WLASL-2000 | 2,000 | 21,083 | 10.5 | 119 Unknown | Scraped | ✗ |
| **ASL Citizen** | **2,731** | **83,399** | **30.5** | **52 Deaf/HH** | **Crowd** | ✓ |

**Figure 3.1:Properties of existing ISLR datasets for ASL compared to ASL Citizen**

### 3.2.1 Data Pre-processing

The original videos of the ASL Citizen Dataset were initially split into three sets for the development of the model. The dataset was split into 70%, 15%, and 15% for training, validation, and testing, respectively, to ensure that the model was adequately trained without overfitting as shown in figure 3.2.

- **Training set:** 500 folders representing different words, each containing approximately **20–21 videos**.
- **Validation set:** 500 folders representing the same words, each containing approximately **6–7 videos**.
- **Test set:** 500 folders representing the same words, each containing approximately **6–7 videos**.

After splitting the dataset, a landmark extraction process was applied to each video using the **MediaPipe** framework. For every frame in each video, keypoints describing the upper body and hand movements were extracted to represent the sign gesture,The extracted landmarks include:

- **Pose landmarks:** 15 selected keypoints from the upper body
  - Points **0–6** correspond to the **nose and eyes**.
  - Points **9–10** correspond to the **mouth corners**.
  - Points **11–16** correspond to the **shoulders, elbows, and wrists**.
- **Hand landmarks:**
  - **Left hand:** 21 keypoints
  - **Right hand:** 21 keypoints

In order to improve robustness, the detected landmarks were further **normalized** by taking the midpoint between the shoulders as a reference and using the distance between the shoulders as a scaling factor. This helps reduce the impact of camera distance and signer position variations.

To preserve the temporal structure of gestures, frames were sampled at **25 frames per second (FPS)** from each video. For every processed video, the extracted landmarks were stored in a **JSON file** containing the video name, label, frame rate, number of frames, and the sequence of pose and hand landmarks.

This preprocessing step converts raw videos into **structured temporal landmark sequences**, which reduces computational complexity while preserving the essential spatial and motion information required for training sequence-based models such as LSTM networks for sign language recognition.

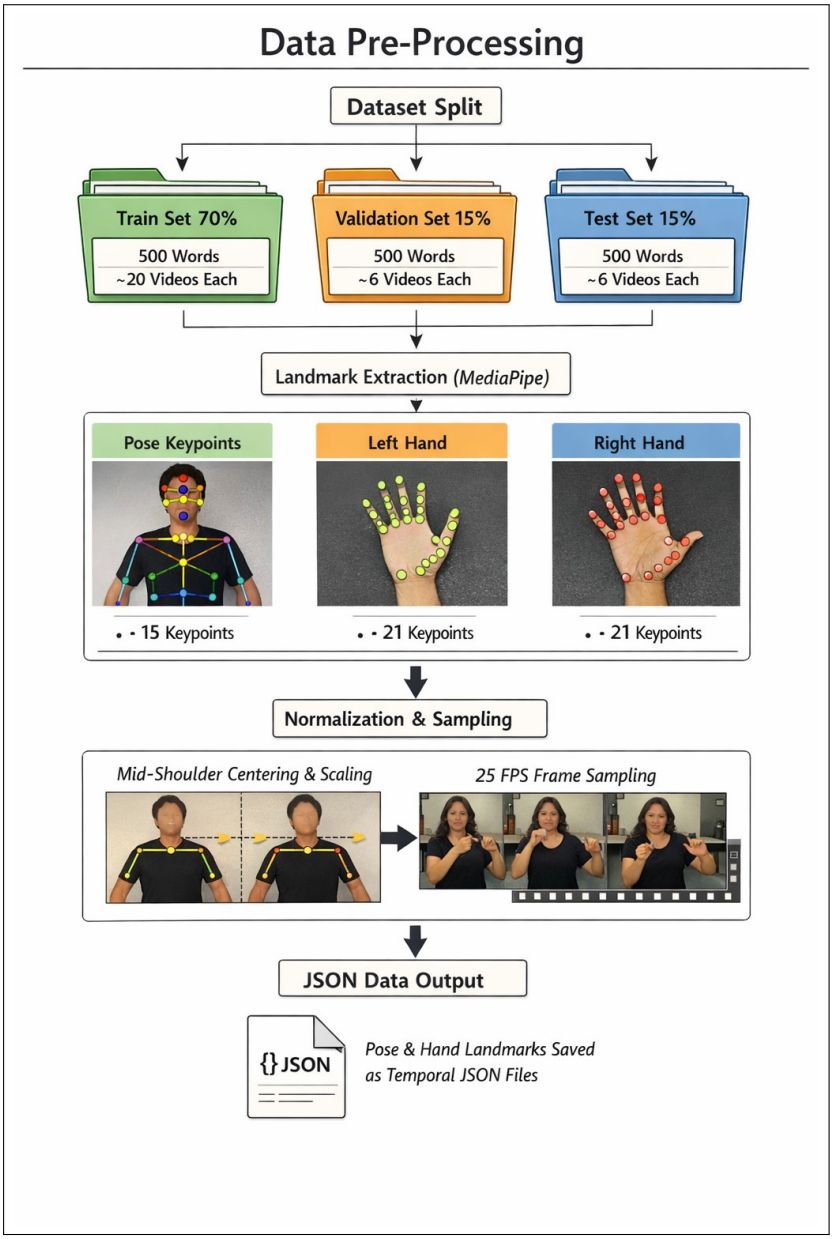



**Figure 3.2:Sign Language Data Pre-processing Workflow**

## 3.3 General architecture of the application

The proposed system is designed to enable effective bidirectional communication between Deaf and hearing individuals through the integration of sign language recognition and natural language processing. The system is intended to enable real-time interaction through an intelligent and cohesive framework.

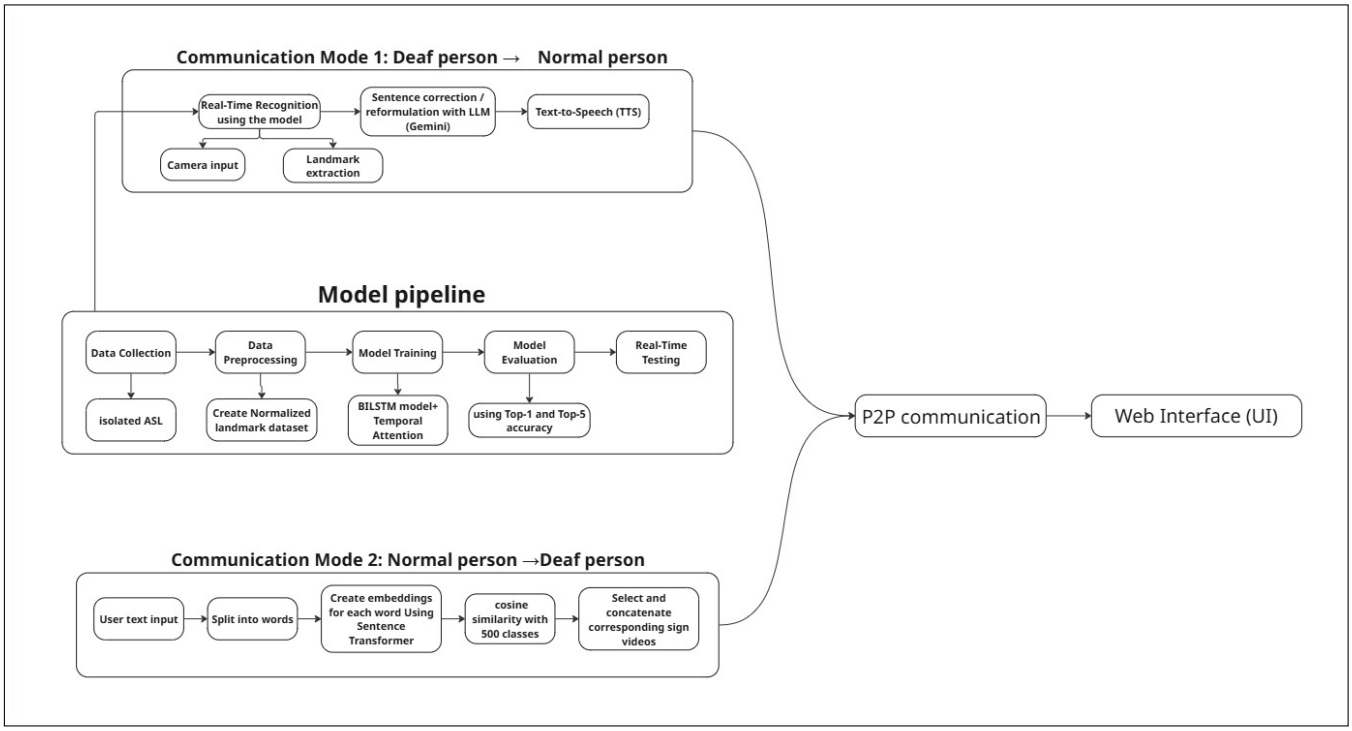



**Figure 3.3:General architecture of the application**

The proposed system architecture as shown in figure 3.3 is structured into four main components: the model pipeline, which handles data processing and prediction , the communication modes (Mode 1 and Mode 2), which define the interaction scenarios, the peer-to-peer (P2P) communication module , which enables real-time data exchange, user interface .

In the following sections, each component is described in detail to illustrate the overall functioning of the system and how these components work together to provide an efficient and practical communication solution.

### 3.3.1 Model Pipeline

The model pipeline is the core component of the proposed system, it is responsible for transforming the input raw data into meaningful predictions. Following the preprocessing stage, which standardizes and prepares the data to enhance learning efficiency, the pipeline integrates the model architecture with training and evaluation procedures. Together, these processes ensure robust, accurate, and reliable performance in real-time sign language recognition.

To successfully model the sequential characteristics of the data, the proposed method uses a **bidirectional long-short-term memory (BiLSTM)** network. This is because this architecture was specifically developed to handle sequential data by processing data in both forward and backward directions, effectively captures past and future dependencies in the data, which is crucial for sign language recognition given the importance of the entire sequence rather than individual frames,based on the following formula:

BiLSTM Formula

$$\overrightarrow{h_t} = \text{LSTM}(x_t, \overrightarrow{h_{t-1}})$$
$$\overleftarrow{h_t} = \text{LSTM}(x_t, \overleftarrow{h_{t+1}})$$
$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$$

**Where:**

- $x_t$ : input at time step $t$
- $\overrightarrow{h_t}$ : forward hidden state at time step $t$
- $\overleftarrow{h_t}$ : backward hidden state at time step $t$
- $h_t$ : concatenated hidden state $[\overrightarrow{h_t}; \overleftarrow{h_t}]$

To further improve the model's ability to focus on relevant information, a **Temporal Attention mechanism** is employed on the outputs of the BiLSTM. The rationale for this is based on the fact that all frames are not created equal when it comes to providing information for gesture semantics. The weights for all frames are determined based on their relative importance, allowing for more relevant information to be highlighted while downplaying irrelevant information,based on the following formula:

Temporal Attention Formula

$$e_t = \tanh(Wh_t + b), \alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)}, c = \sum_{t=1}^{T} \alpha_t \cdot h_t$$

**Where:**

- $h_t$ : hidden state at time step $t$ from LSTM/BiLSTM
- $e_t$ : attention score for time step $t$
- $\alpha_t$ : attention weight for time step $t$
- $c$ : context vector computed as weighted sum of hidden states
- $W, b$ : learnable parameters of the attention layer

The model is trained on pre-defined data; this data provides a sequence of actions and their meanings. During the model's training, the order in which the actions occur and how they are related are determined. We measure its accuracy using the Top-1 and Top-5 criteria, which show us how well the model predicts and, more importantly, its ability to understand the sequence of hand movements as they are performed.

### 3.3.2 Communication Modes

The system operates through two distinct communication modes, each designed to address specific interaction scenarios between deaf and hearing users. These modes ensure bidirectional communication, allowing both parties to understand each other in real time.

#### 3.3.2.1 Communication Mode 1

This mode begins with real-time video input captured through a camera, where the deaf user performs sign language gestures. To ensure consistency between training and inference, the incoming video stream undergoes the same preprocessing steps applied during dataset preparation.

Specifically, each frame is processed to extract spatial landmarks representing key points of the hands, uper body (15 point ). These landmarks provide a compact and robust representation of the gesture, reducing sensitivity to background noise, lighting variations, and camera conditions. The extracted landmark sequences are also normalized in the same manner as the training data, ensuring compatibility with the trained model.

This modality uses a sign language recognition model, which has been developed by applying the previously discussed model pipeline and has been trained on an isolated sign language dataset. As a result, it produces a list of isolated labels or words, which match the signs.

Generally, the output of the sign language recognition model will consist of a list of isolated words, for instance, "Teacher," "Go," "School." This output indicates the meaning of the signs performed, but it does not include grammatical correctness, syntactic correctness, and contextual correctness, which are essential for natural language,for this purpose, a post-processing layer has been incorporated into the system, which uses the state-of-the-art Large Language Model (LLM), named Gemini. This layer will take the list of isolated words as input and will convert it into a grammatically correct sentence.

Furthermore, the system includes an optional text-to-speech (TTS) module. This module will convert sentences into audible speech for the hearing user, thus expanding communication options.

For this purpose, in this study, we used the Gemini 2.5 Flash for sentence formation because it is efficient for real-time applications[68]. It has a constraint of 20 requests per day (RPD), which is sufficient for experimental purposes.

### 3.3.2.2 Communication Mode 2

This mode helps the hearing user communicate with the Deaf user by converting the text sentence into a video sequence of signs in the sign language,this mode starts with the text provided by the user. The text sentence goes through the preprocessing step in which the sentence is divided into words and punctuation symbols are removed.

Next, the words in the sentence are converted into numerical form using the concept of "Semantic Embeddings" provided by the "Sentence Transformers" model. Semantic Embeddings transform the semantic content of the text data into numerical vectors. These vectors help the model in calculating the similarity between the words.

In this study,, the pre-trained model "all-MiniLM-L6-v2" has been used. The model provides 384-dimensional vectors. These vectors are more efficient in calculating the similarity between the words. The "Sentence Transformers" model has the advantage over the "Word Embeddings" model, such as "Word2Vec" and "GloVe," in the sense that the "Sentence Transformers" model has the capability to understand the sentence-level semantics.

Finally, the similarity between the words in the sentence and the known signs in the sign language is calculated by finding the cosine similarity between the words in the sentence and the database of 500 classes and select the top 1 ,based on the following formula:

Cosine Similarity Formula

$$\text{cos_sim}(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \, \|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

**Where:**

- $\mathbf{A}, \mathbf{B}$ : embedding vectors (e.g., word or sentence embeddings)
- $\mathbf{A} \cdot \mathbf{B}$ : dot product between vectors $\mathbf{A}$ and $\mathbf{B}$
- $\|\mathbf{A}\|, \|\mathbf{B}\|$ : Euclidean norms (magnitudes) of the vectors
- $A_i, B_i$ : components of the vectors in dimension $i$
- $n$ : dimensionality of the embedding space (e.g., 384)
- $\text{cos_sim} \in [-1, 1]$ : similarity score (closer to 1 indicates higher similarity)

A decision mechanism is then employed based on three specific conditions:

1. **Coverage condition:** At least 80% of the words in the sentence to be processed must have a valid match with the predefined classes.
2. **Word-level similarity condition:**The similarity measure between each word and the class to which it is matched must be greater than 70%.
3. **Sentence-level similarity condition:**Similarly, the cosine similarity measure between the sentence and the class match words must be greater than 60%.

Should all the prerequisites be met, then the system will fetch all the videos corresponding to each word match. The resultant videos will be put into some form of sequence to make up a seamless video representing the sentence.

In cases where all the prerequisites are not met, the behavior of the system would revert to the normal one where the sentence is sent out only in text form. This way, communication is still ensured.

### 3.3.3 Peer-to-Peer Communication

The system uses the P2P (peer-to-peer) communication model that helps to implement direct communication between peers without involving the central server. Under this approach, every participant acts as both client and server, which allows for direct transmission of messages from sender to receiver. The approach is suitable for real-time systems because it increases speed while avoiding additional intermediate processors.

There are several reasons why P2P communication was chosen. First, it increases efficiency of the process because the delays related to using servers are minimized. Second, it increases scalability of the communication process as it does not require involvement of the central server that might turn into bottleneck. Third, it provides better privacy as private information is exchanged directly between participants.

However, the lack of any central controller creates additional security concerns, especially, in terms of data interception, data manipulation, and unauthorized access. Therefore, it is essential to provide effective solutions to guarantee secure data exchange between participants.

Security of the communication process implemented within the P2P network can be guaranteed via confidentiality, integrity, and authentication procedures. All measures mentioned above perform different roles:

1. **Confidentiality – Hybrid Encryption (RSA + AES-GCM):**To ensure that only the intended recipient can read the message, the system encrypts the plaintext message $m$ using AES-GCM, a symmetric encryption algorithm. The symmetric key $K$ is then encrypted using the recipient's RSA public key $pk$. This guarantees that even if the communication is intercepted, the message content remains confidential.
2. **Integrity – AES-GCM Authentication:**AES-GCM provides an authentication tag that allows the recipient to verify that the message has not been modified during transmission. Any alteration in the ciphertext $C$ will cause the decryption to fail. This ensures the integrity of the transmitted message.
3. **Authentication:**Each message is signed by the sender using their RSA private key $sk$ via RSA-PSS. The recipient can verify this signature using the sender's public key $pk$. This guarantees that the message indeed originates from the claimed sender and has not been forged.
4. **Signature Verification:**Upon receiving the message, the recipient checks the digital signature $\sigma$. If the verification succeeds, the message is trusted; otherwise, it is discarded. This step ensures both authentication and integrity.

For reliable end-to-end transmission, each video chunk is associated with a hash value for verification.If a mismatch is detected, indicating a lost or corrupted chunk, the system automatically resends the missing chunks, ensuring accurate reconstruction of the video.

### 3.3.4 User Interface

The interface for the system is implemented as a web application, using HTML, CSS, and JavaScript on the front end, complemented by a Python Flask back end to facilitate live interaction. The roles of the users are divided into three types: normal (hearing) users, deaf users, and administrators, each assigned an interface based on their respective roles. The input text submitted by normal users is analyzed by the back end to produce videos in sign language. The deaf users interact through live video input, where the gestures are analyzed and converted into text and optionally audio. The administrators use a dashboard interface to track interactions and manage users.

## 3.4 System Design With Uml

### 3.4.1 UML

UML (Unified Modeling Language) is a standardized modeling language used in software engineering and system design. It provides a set of graphical notations to represent various aspects of a system, including its structure, behavior, and interactions. UML diagrams help communicate complex ideas and concepts among stakeholders during the software development process[69].

- **Type of diagrams**

1. **Use Case Diagram**:
   - Depicts interactions between **actors** (users or external systems) and **use cases**.
   - Shows **functional requirements** from a user's perspective.
   - Useful for **requirements analysis** and **system design**.
2. **Class Diagram**:
   - Represents the **static structure** of a system.
   - Shows **classes**, their **attributes**, **methods**, and **relationships**.
   - Useful for understanding **the class hierarchy** and **inter dependencies**.
3. **Sequence Diagram**:
   - Displays **object interactions** over **time**.
   - Shows the **sequence of messages** exchanged between objects.
4. **Activity Diagram**:
   - Visualizes **workflows**, **processes**, and **business logic**.
   - Includes **activities**, **control flow**, and **decision points**.

### 3.4.2 Use Case diagrams

The use case diagram shows the functionalities provided by the system through several actors, in our case they are the normal person and the deaf/non verbal person .

- **Normal person use case diagram:** is presented in the following diagram in the figure 3.4

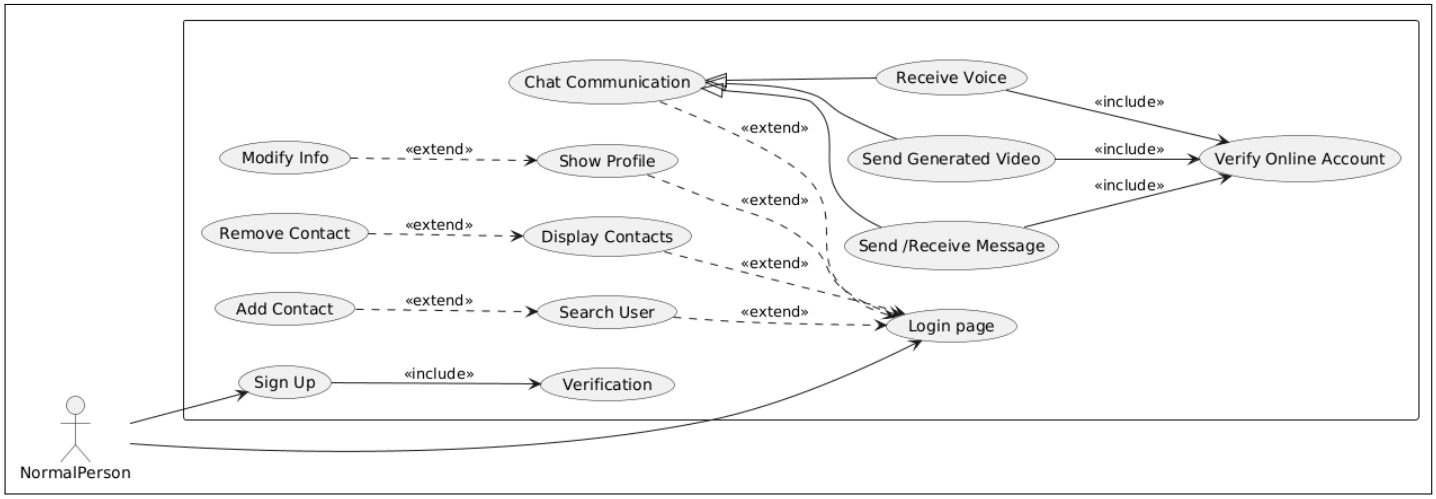



**Figure 3.4:normal person use case diagram**

- **Deaf/Non Verbal person use case diagram:** is presented in the following diagram in the figure3.5

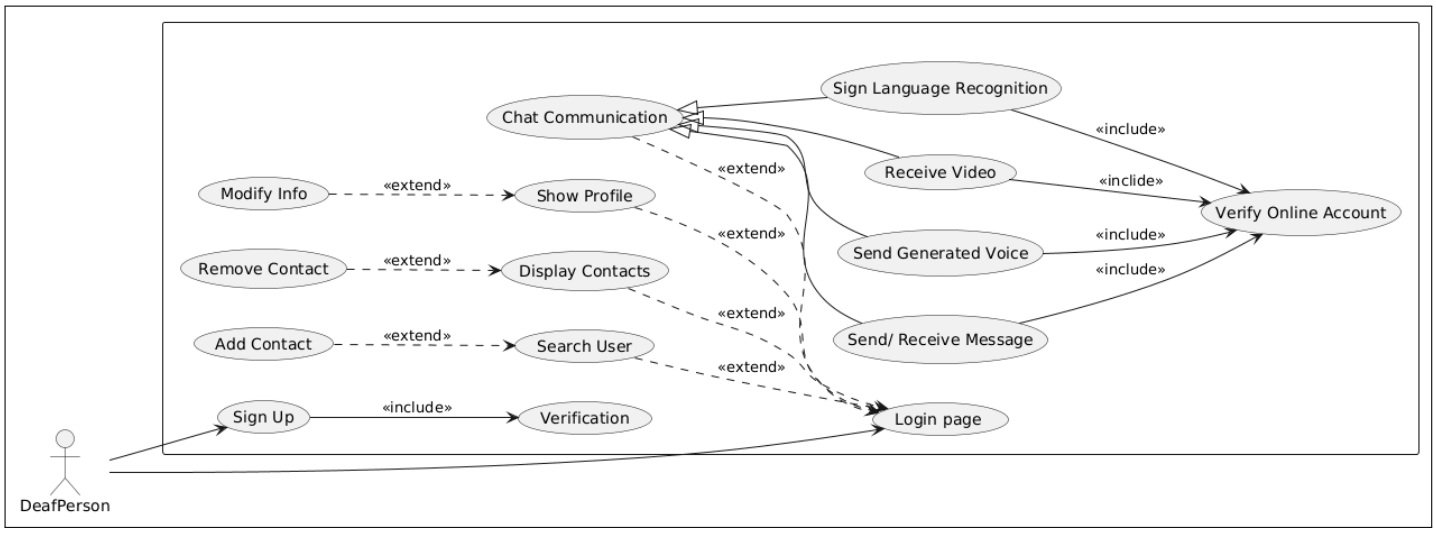



**Figure 3.5:deaf person use case diagram**

### 3.4.3 Class diagrams

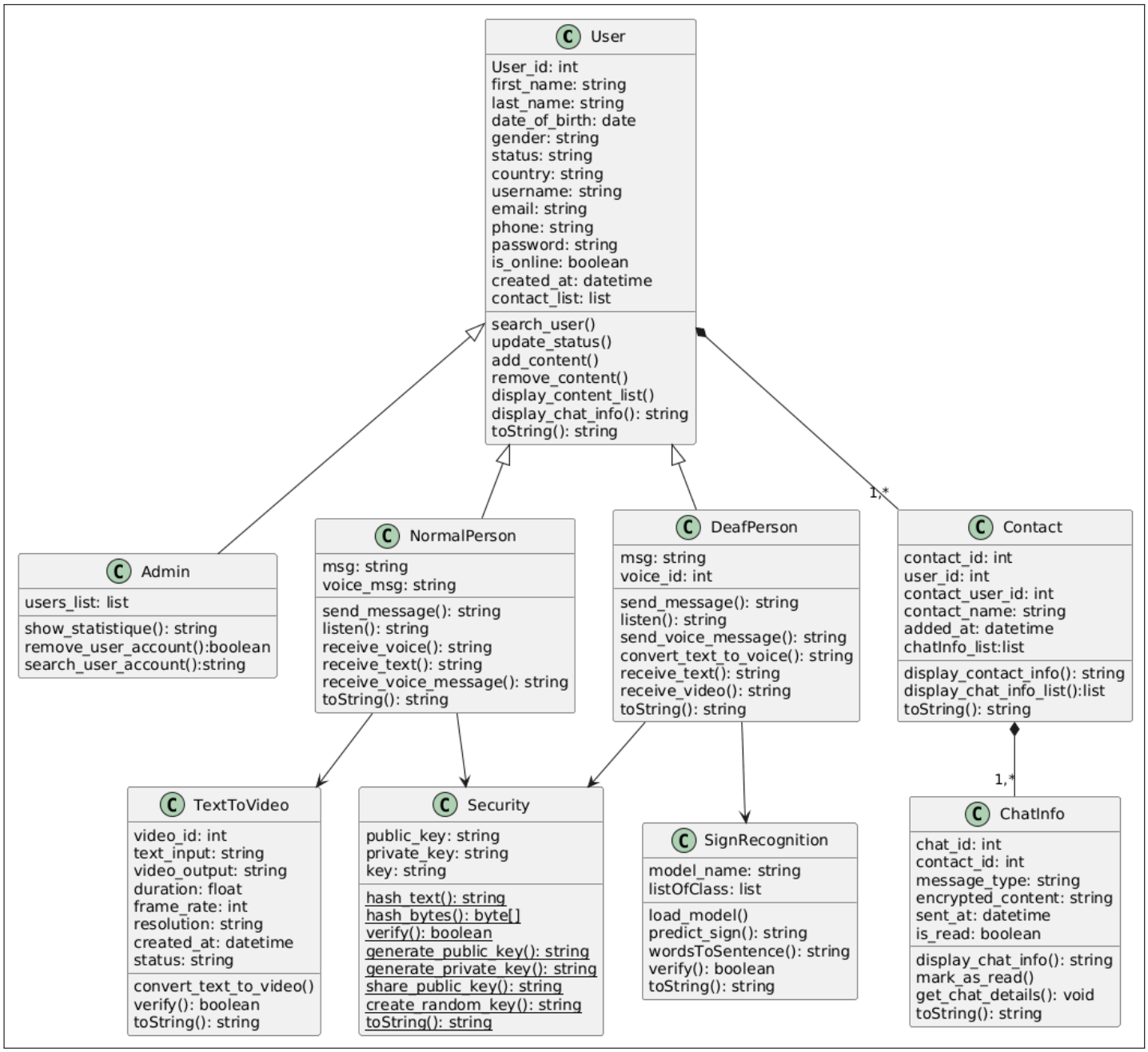



**Figure 3.6:class diagram**

### 3.4.4 Sequence diagrams

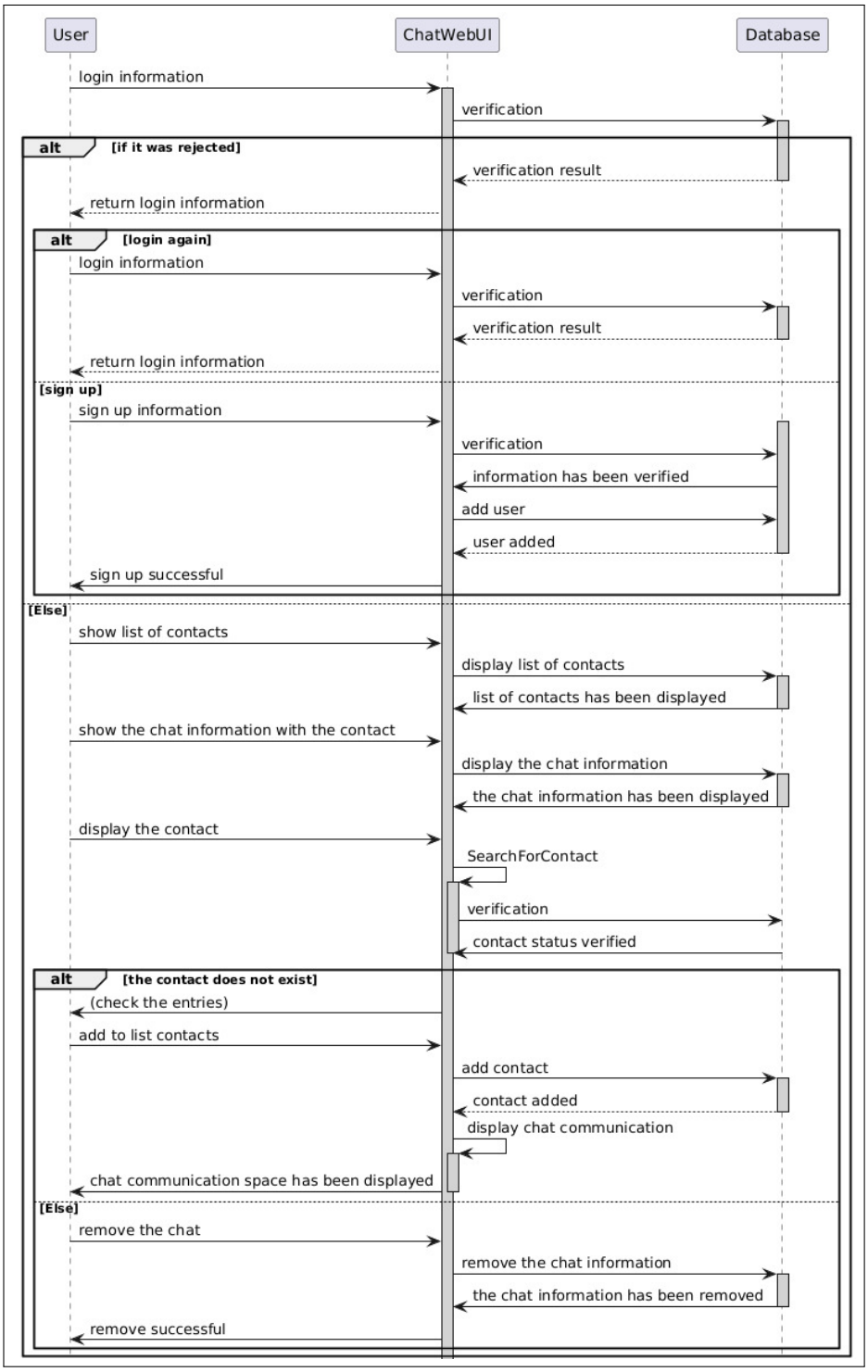



**Figure 3.7:User sequence diagram**

### 3.4.5 Activity diagrams

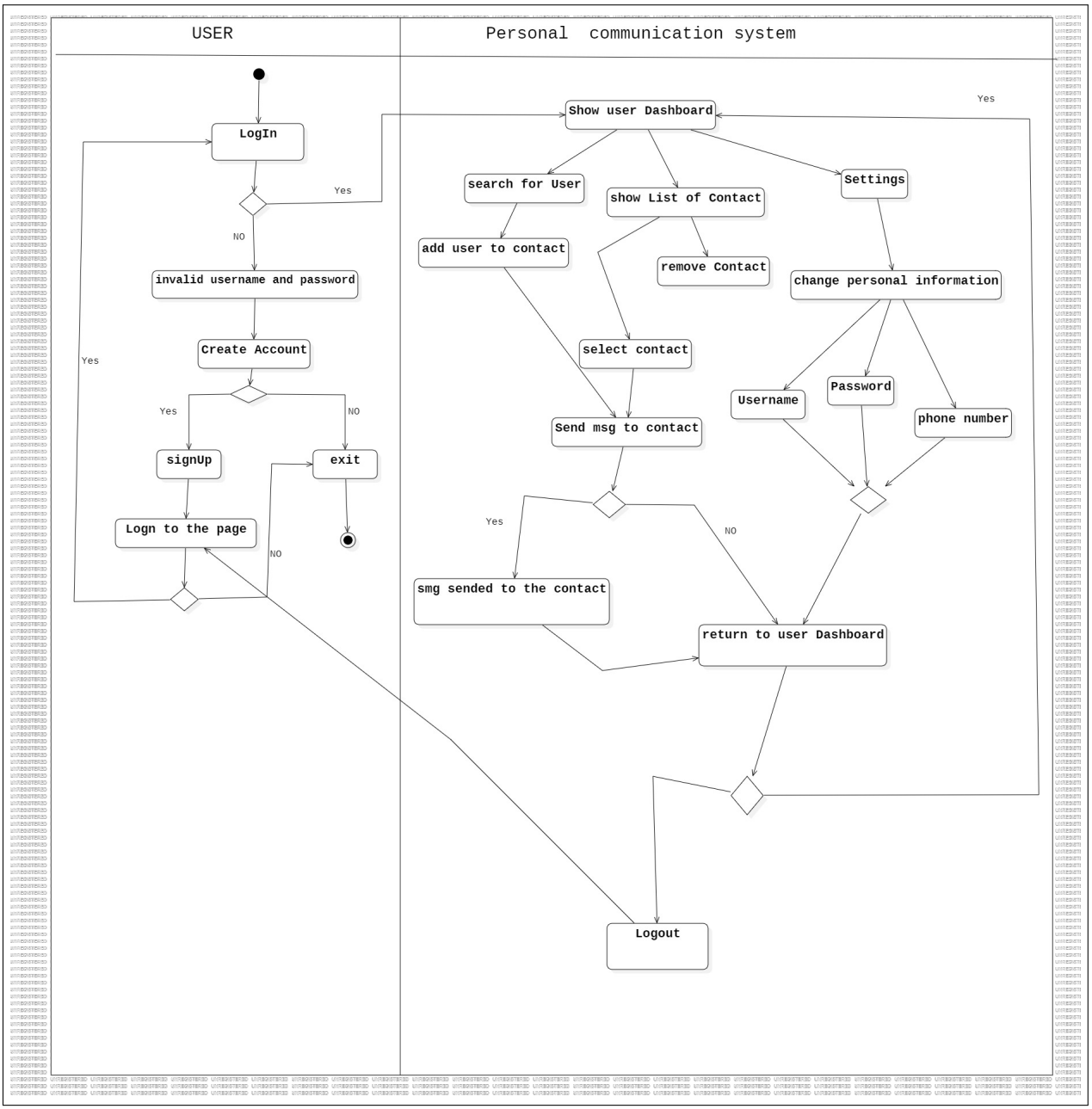



**Figure 3.8:User activity diagram**

# 3.5 Database design

## 3.5.1 Association entity model

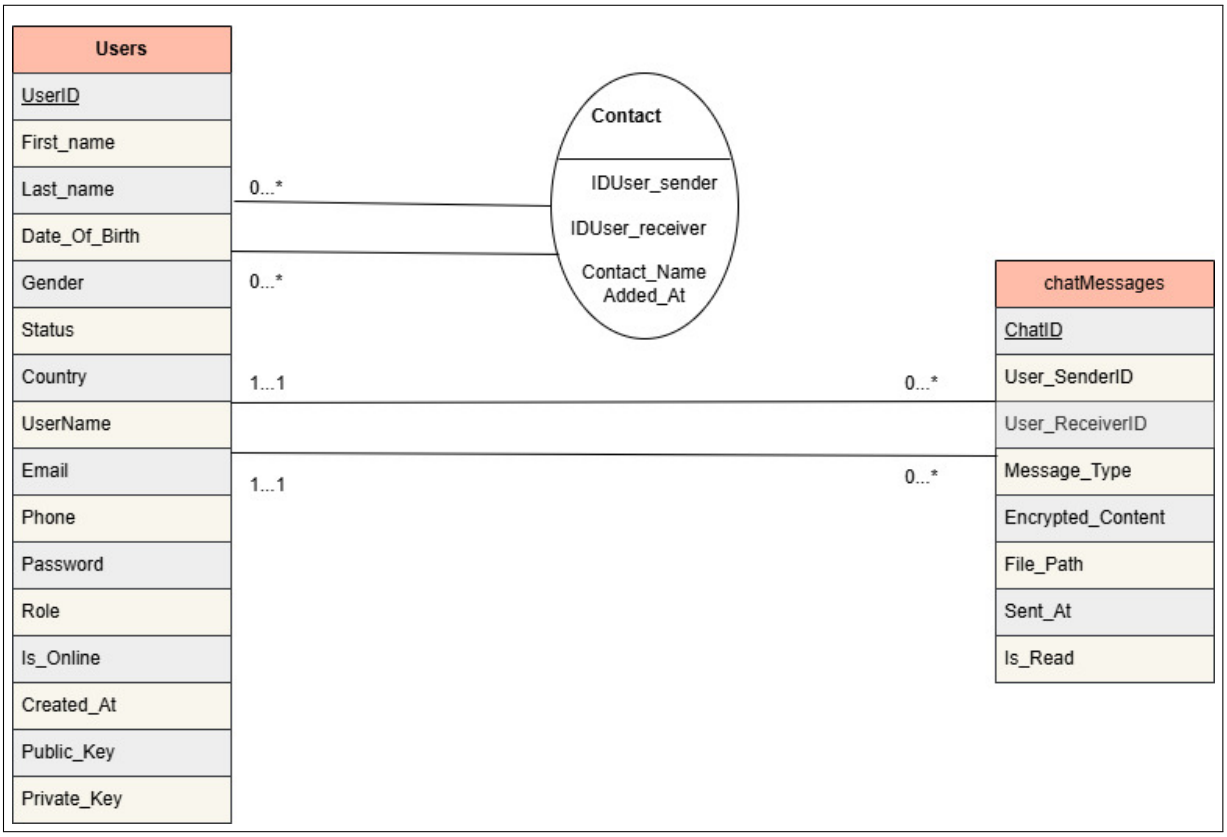



**Figure 3.9:Database association entity model**

## 3.5.2 Entity–relationship model

- Users ( user_id , first_name , last_name , date_of_birth , gender , status , country , username , email , phone , password , role , is_online , created_at , public_key ,private_key )
- Contact ( contact_id , #user_sender_id , #user_receiver_id , contact_name , added_at )
- ChatMessages ( chat_message_id ,#user_sender_id , #user_receiver_id , message_type , encrypted_content , file_path, sent_at , is_read )

# 3.6 Conclustion

This section focuses on the design and implementation of web application for communication between deaf/non-verbal and hearing person. This section describes dataset, architecture of the system, design of system in terms of UML and entity-relationship diagrams,which will help determine the optimal programming languages for designing the user interface. The next chapter focuses on the implementation environment, accuracy of the model, user interface of the application.

# Chapter4- Implementation and Results

# 4.1 Introduction

In most systems of sign language recognition, their main goal is to establish a connection between deaf people and those who have impaired hearing capabilities by translating input gestures into either text or speech. While there have been advancements in areas such as computer vision and deep learning, the development of an application whose integration of both hardware and software components will deliver high-performing models and an efficient user interface still poses challenges.

In designing sign language recognition systems, developers must consider the hardware components needed to efficiently collect and process data, and appropriate software technologies for implementing the systems. Also, factors such as datasets, training techniques, and variability in users and environments impact the level of performance achieved by developed models. Finally, designing an efficient user interface requires that consideration be accorded to usability issues.

This chapter provides the implementation of the proposed application in terms of hardware and software used, model specification and results achieved, as well as user interface setup.

# 4.2 Development environment

To implement our application , we utilized a personal computer with the following specifications:

| **Component** | **Specification** |
|---|---|
| Processor | Intel Core i5-12400 (6 cores, up to 4.4 GHz) |
| RAM | 16 GB DDR4 |
| GPU | NVIDIA GeForce RTX 3060 Ti (8 GB GDDR6) |
| System Type | 64-bit Operating System, x64-based processor |
| Edition | Windows 10 Pro |

Table 4.1 :Characteristics of the used material

## 4.2.1 Programming language

In this study, we used the programming languages HTML, CSS, JS and Python with a MySQL database,which will be detailed as follows:

### 4.2.1.1 Python

It is an interpreted, high-level programming language noted for its easy readability. Created by Guido van Rossum, it was first released in 1991. Python is a versatile language that can be used in different programming paradigms like procedural, object-oriented, and functional. As a result, Python has a wide range of uses and can be applied to many fields, for instance, website creation and artificial intelligence[70].

#### 4.2.1.2 HTML(Hyper Text Markup Language)

is the most basic building block of the Web. It defines the meaning and structure of web content. Other technologies besides HTML are generally used to describe a web page's appearance or presentation(CSS) or functionality/ behavior(JavaScript)."hypertext" refers to links that connect web pages to one another, either within a single website or between websites[71].

#### 4.2.1.3 CSS(Cascading Style Sheets)

CSS is a language used to describe the presentation of a document written in HTML. It enables you to create visually appealing web pages by controlling the layout, colors, fonts, and overall design. It works with HTML to create the formatting of web pages, giving websites a sleek and professional look[72].

#### 4.2.1.4 JavaScript

JavaScript is a powerful scripting language that is used in developing highly interactive web pages. It facilitates easy user interaction, dynamic updates, and advanced web applications. Being one of the essential technologies in creating web pages together with HTML and CSS, JS works with any popular browser and allows for adding interactive elements, handling multimedia, and animating graphics on websites. The uses of JS range from simple tasks to elaborate online games and enterprise applications[73].

#### 4.2.1.5 MySQL

is an open source relational database management system that we caninteract with using structured query language (SQL). Query language is a simple programming language that can help you access, process and manage data in a relational database. We use MySQL to store all the information you need to run large and complex websites efficiently[74].

- Development platform:

1. XAMPP:XAMPP is the address used to bundle free software. The name is an abbreviation of (X Apache Mysql Php Perl), where each letter represents one of the five main components. The software package contains the Apache web server, the MySQL (or MariaDB) relational database management system, and the Perl and PHP scripting languages. The initial X indicates the operating systems it works with: Linux, Windows, and Mac OS[75].

- Management of Data Base:

1. PhpMyAdmen:It is a free software tool written in PHP, intended to handle MySQL administration over the web. phpMyAdmin supports a wide range of operations on MySQL and MariaDB[76].

### 4.2.2 Libraries

In this study, we use several libraries, which will be detailed as follows:

#### 4.2.2.1 Pyhorch

is an open source machine learning framework that accelerates the path from research prototyping to production deployment. Built to offer maximum flexibility and speed, PyTorch supports dynamic computation graphs, enabling researchers and developers to iterate quickly and intuitively. Its Pythonic design and deep integration with native Python tools make it an accessible and powerful platform for building and training deep learning models at scale[77].

#### 4.2.2.2 MediaPipe

MediaPipe is a cross-platform framework designed for building applied multimedia machine learning pipelines. It allows developers to create complex systems for processing time-series data such as video and audio with a focus on real-time applications, and is widely used for tasks such as hand and face detection, object tracking, and more[78].

#### 4.2.2.3 OpenCV

is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being an Apache 2 licensed product, OpenCV makes it easy for businesses to utilize and modify the code[79].

#### 4.2.2.4 Flask

is a Python web micro-framework distinguished by its lightweight nature and flexibility. Created in 2010 by Armin Ronacher as part of the Pallets project, Flask was born from an April Fool's joke that turned into a serious project. The term "micro" doesn't mean Flask is limited in features, but rather that it keeps a simple, extensible core. Flask makes no assumptions about the database, template system (though it integrates Jinja2 by default), or authentication layer you'll use[80].

#### 4.2.2.5 MoviePy

is a Python module for video editing, which can be used for basic operations (like cuts, concatenations, title insertions), video compositing (a.k.a. non-linear editing), video processing, or to create advanced effects. It can read and write the most common video formats, including GIF[81].

#### 4.2.2.6 Cryptography

is a Python module that provides constructs along with basic cryptography primitives to provide security for the encryption and decryption process. This helps developers to build applications with strong security features[82].

#### 4.2.2.7 BcryptJS

is a JavaScript package used to implement bcrypt hash algorithms for secure password hashing and validation purposes.been popularly used in web-based applications to provide authentication services and strengthen password protection[83].

#### 4.2.2.8 Socket (Python)

is a way of connecting two nodes on a network to communicate with each other. One socket(node) listens on a particular port at an IP, while the other socket reaches out to the other to form a connection. The server forms the listener socket while the client reaches out to the server[84].

## 4.3 Model Performance and Analysis

The following is an illustration of how the results have been obtained in this work, which consist of three parts: training details, model results and analysis , and model discussion. It is important to highlight that isolated words are used to increase the ability to generate flexible sentences (500 words).

### 4.3.1 Training Details

For training purposes, we used the Adam optimizer which allows dynamically modifying the learning rate in order to enable efficient convergence. We started with a learning rate of 0.001, and the total number of training epochs was 100, while the batch size was equal to 32. For modeling the data, a bidirectional LSTM network with two layers and an attention layer that allows capturing temporal dependencies in the input sequence was used. In order to prevent overfitting, we have used a dropout rate of 0.3 in LSTM layers. For calculating the loss, we have used cross-entropy loss because the problem was a multi-class classification, where an error after each prediction is calculated and used for the Adam optimizer to update the weights. Moreover, the issue of varying sequence lengths was solved by padding and packing. Accuracy, Top-1 accuracy, Top-5 accuracy, Precision, Recall and F1-score were chosen as the measures for model evaluation.

### 4.3.2 Model Results and Analysis

Here we present the results of a sign language recognition model. Evaluation of the performance of the model is done by using various metrics that are calculated based on the confusion matrix. The confusion matrix consists of four elements: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Based on these elements, some important performance indicators such as loss [85], accuracy, precision, recall, and F1-score can be calculated[86]:

- **True Positives (TP):** The model correctly predicted a positive outcome i.e the actual outcome was positive.
- **True negatives (TN):**The model correctly predicted a negative outcome i.e the actual outcome was negative.
- **False Positives (FP):**The model incorrectly predicted a positive outcome i.e the actual outcome was negative. It is also known as a Type I error.
- **False Negatives (FN):** The model incorrectly predicted a negative outcome i.e the actual outcome was positive. It is also known as a Type II error.

- **Accuracy:** Accuracy measures the correctness of the results produced by a system or model by dividing the number of correct predictions by the total number of predictions and multiplying it by 100 to get a percentage. High accuracy indicates that a system or model is performing well in its predictions:

$$\boxed{\text{Accuracy (\%)} = \frac{TP + TN}{TP + TN + FP + FN} \times 100}$$

- **Precision:** focus on the quality of the model's positive predictions. It tells us how many of the "positive" predictions were actually correct. It is important in situations where false positives need to be minimized such as detecting spam emails or fraud. The formula of precision is:

$$\boxed{\text{Precision} = \frac{TP}{TP + FP}}$$

- **Recall:**measures how how good the model is at predicting positives. It shows the proportion of true positives detected out of all the actual positive instances. High recall is essential when missing positive cases has significant consequences like in medical tests:

$$\boxed{\text{Recall} = \frac{TP}{TP + FN}}$$

- **F1-Score:**combines precision and recall into a single metric to balance their trade-off. It provides a better sense of a model's overall performance particularly for imbalanced datasets. It is helpful when both false positives and false negatives are important though it assumes precision and recall are equally important but in some situations one might matter more than the other:

$$\boxed{\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}}$$

- **Loss:** A loss function is a mathematical way to measure how good or bad a model's predictions are compared to the actual results. It gives a single number that tells us how far off the predictions are. The smaller the number, the better the model is doing. Loss functions are used to train models.

## • BILSTM model

The results of the BILSTM model presented and detailed in the previous chapter are shown in the Table 4.2:

| Validation Loss | Top-1 Accuracy | Top-5 Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| 1.3147 | 0.7291 | 0.8784 | 0.7364 | 0.7126 | 0.7055 |

Table 4.2: Evaluation results of classification bilstm model

From the results,the model achieved a validation loss rate of approximately 1.3147, indicating a clear mismatch between the predicted classifications and the correct categories due to the multi-category nature of the classification .

Regarding the results, it is worth noting that the model demonstrates a 0.7291 validation accuracy for the Top-1 category and an 0.8784 validation accuracy for the Top-5 category, meaning that the predicted classification belongs to the top five categories with a high probability. Furthermore, the resulting accuracy, recall, and F1 metric values of 0.7364, 0.7126, and 0.7055, respectively, indicate good predictive equilibrium ,although the number of classes was considered a limitation.

Two challenges are particularly noteworthy. The first challenge is the large number of categories (500 American Sign Language signs), which increases the complexity of the classification process and promotes similarity between categories. Second, some American Sign Language signs contain similar movements, making differentiation difficult due to their varying representations.

Overall, the results show that the model is able to learn features efficiently from the provided American Sign Language data, regardless of training difficulties, The accuracy and loss curve are shown in figure 4.1 .

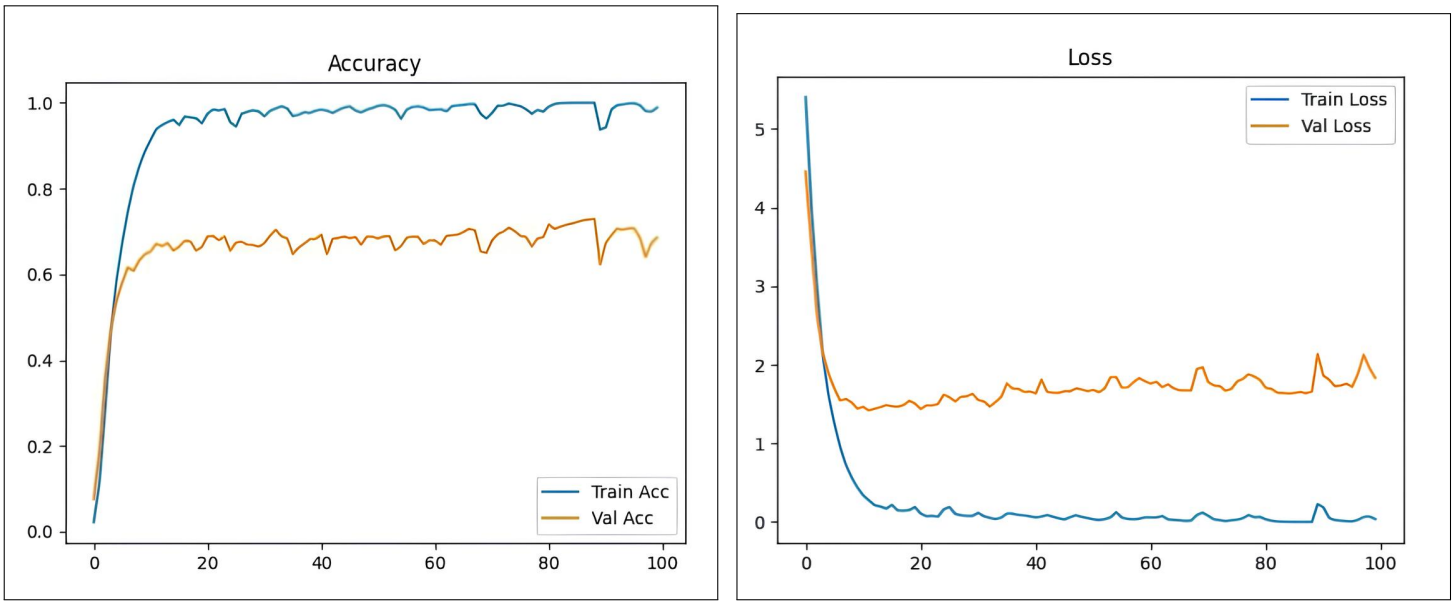



**Figure 4.1:Accuracy and training loss Curve of BiLSTM model.**

the figure 4.2 represents the confusion matrix of the suggested approach. Given the large number of categories, the resulting confusion matrix is quite dense, and confusions occur mostly among semantically similar symbols. nonetheless, high values on the diagonal of the confusion matrix show that the classifier makes correct classifications in many cases:

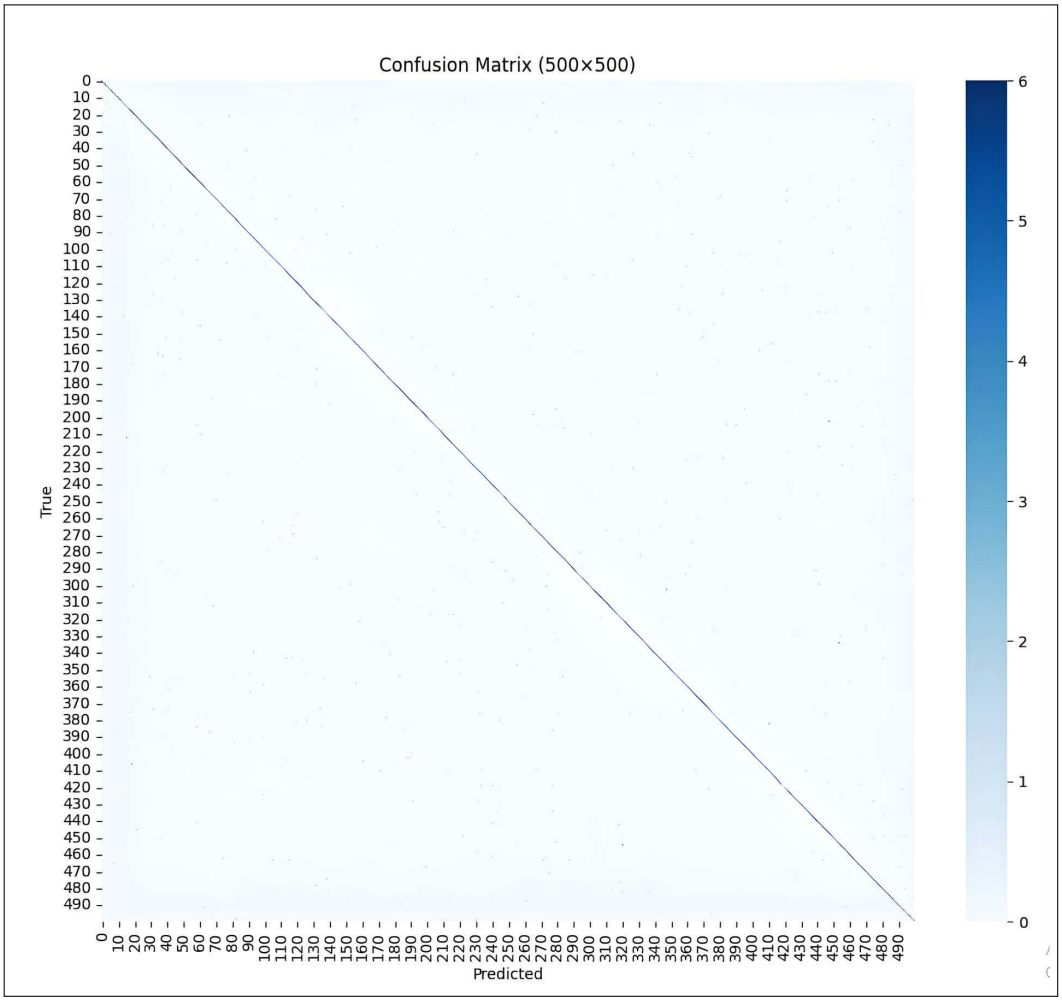



**Figure 4.2:Confusion Matrix of BiLSTM model.**

### 4.3.3 Result discussion

This section discusses the obtained results through two perspectives: result evaluation, which measures system performance, and result comparison, which examines the findings in relation to existing studies and approaches.

#### 4.3.3.1 Result evaluation

According to the obtained results, the developed model shows good performance in recognizing various types of signs in videos. Due to the use of a combined architecture based on BiLSTM and upper-body features, which include 15 key points, the model is able to detect both spatial and temporal dependencies in gestures and their movements. This approach solves some major challenges in sign language recognition tasks.

Main strengths and contributions of the designed system can be described as follows:

1. Designing of a new approach, where BiLSTM network and attention mechanism work together to process sequence-based data (not images).

2. The use of only upper-body characteristics (15 points). In other words, the input dimensions are significantly reduced while preserving important movement characteristics.

3. The model's ability to detect relationships between signals in a sequence and improve dynamic gesture recognition performance.

4. Recognition performance remains good regardless of gesture variety, demonstrating the model's generality across different samples.

5. Relationships between signals in sequences are considered to enhance dynamic gesture recognition.

6. . The system's good performance, regardless of gesture variety, demonstrates its generality across diverse samples.

   This approach relies primarily on representing input data using a video stream, where time sequences are more important than still images,its promising, especially for classifying large video sets containing dynamic signals belonging to multiple categories.

#### 4.3.3.2 Result comparison

In trying to compare our results with previous researches, we have met some problems since there is a lack of articles that used a similar dataset and a comparable amount of classes. most of the relevant papers make use of specially created or manually constructed datasets for particular experimental conditions and only a few classes. therefore, we had to choose the closest and the most relevant papers for comparison purposes,as shown in Table 2.1.

Unlike previous papers that used a limited amount of classes, the presented model with the help of BiLSTM structure, upper-body features (15 points), and temporal attention succeeded in recognizing signs within the ASL Citizen dataset with 500 classes.

The following problems appeared in developing the system:

Firstly, the major problem was the high cost of deep learning algorithms' training on the video dataset. one clip of the video is from 600 KB to 1 MB, and consequently, training is expensive because of consuming much memory and computing resources, particularly, because of a large number of classes. Moreover, it is required to achieve the best possible results without overfitting.

Additionally, issues arose related to using different library versions, which impacted training stability. Finally, the selection of a video dataset aligned with the project's objectives was critical. Most of the existing video datasets were filmed in an environment designed to achieve the desired results, and our goal was to find a video dataset that closely resembled a real-world environment.

However, despite the described difficulties, the system was adjusted to the ASL Citizen dataset of 500 sign languages classes.

## 4.4 Overview of Application Interface

Any user who accesses our web application, whether an administrator, a regular person, or a deaf person, will be taken to the login page or the registration page:

### 4.4.1 Login

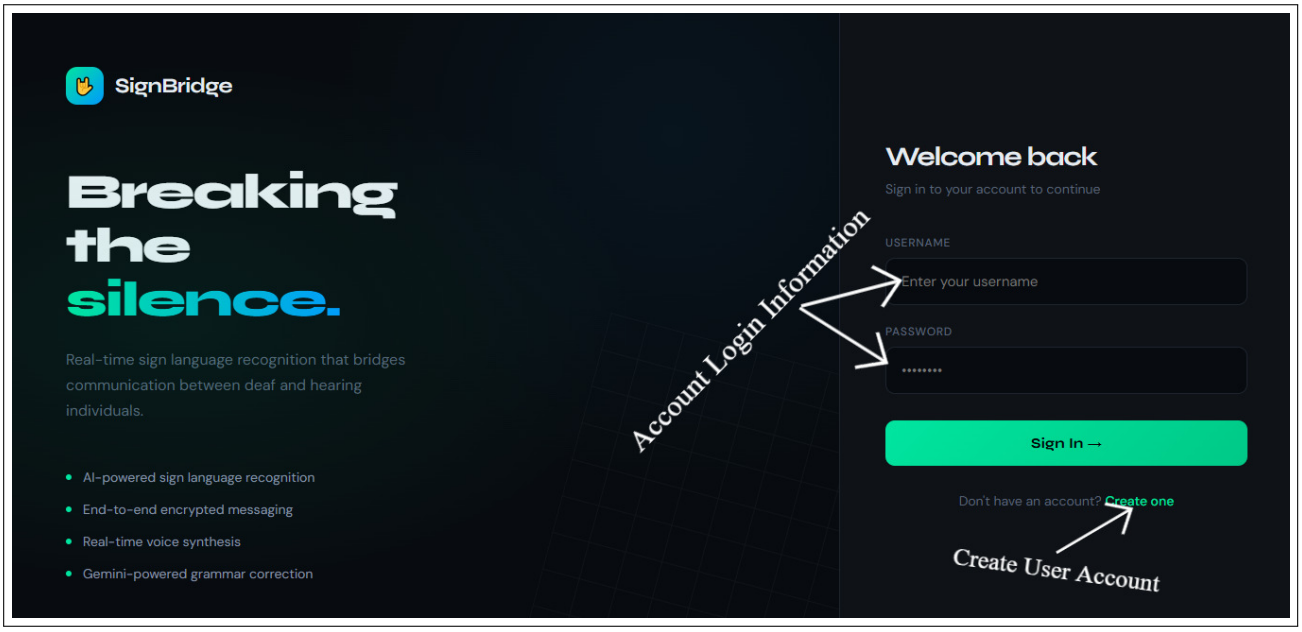



**Figure 4.3:Login page view.**

In order for the user to get into the chat area, the user first have to perform authentication by logging in the user account. The login page is used to make sure that only the allowed users may access the system as can be observed from the figure 4.3, personal information security and user's privacy would not be compromised. Moreover, authentication allows the system to distinguish who the user is so to allow user to communicate safely.

### 4.4.2 Sign-Up

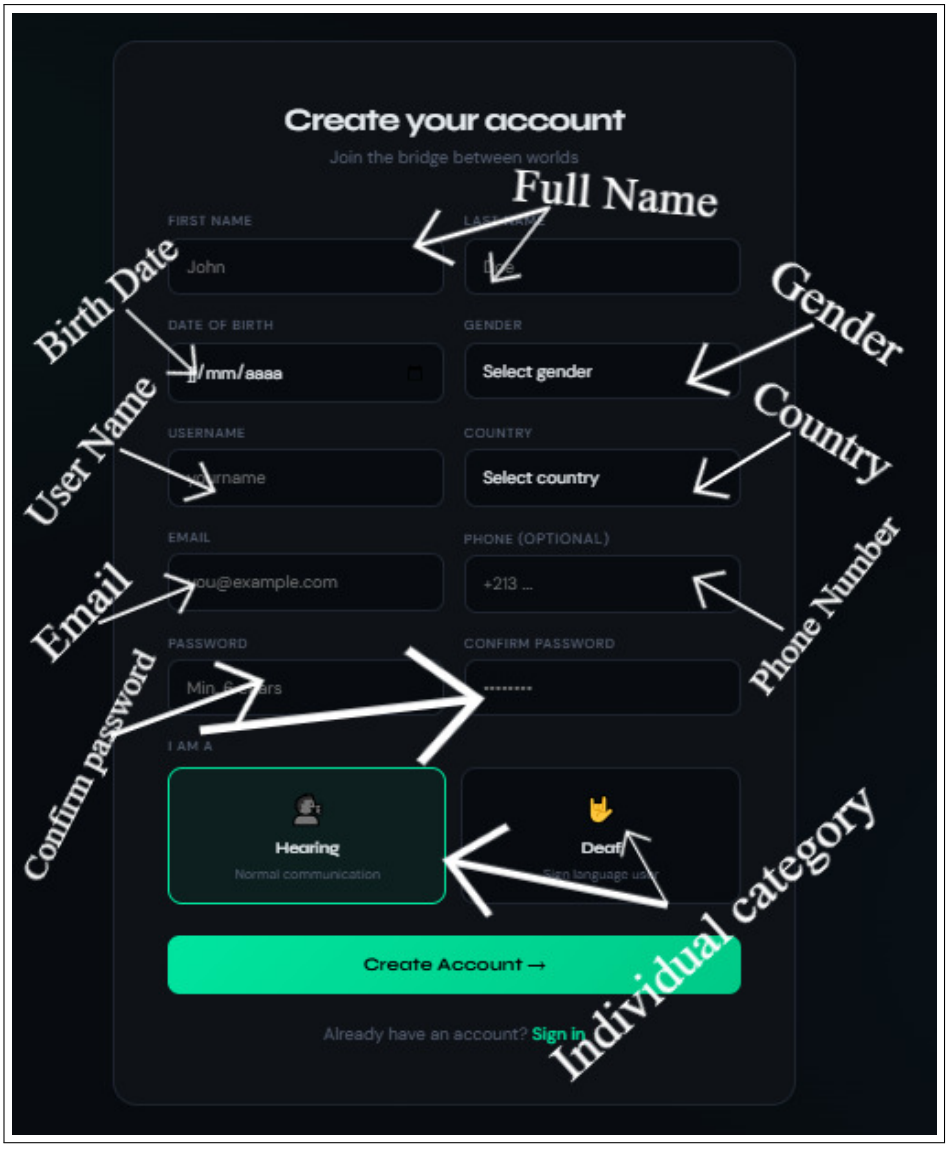



**Figure 4.4:Sign-Up page view.**

With signUp page, new users can make themselves an account and get to use the communication system in the platform. In the page, users will need to add details like: First name, Last name, Country, phone number, E-mail, and Password. Apart from those, the users have to choose what kind of person are they, either normal or deaf user, also, gender,as shown in figure 4.4.

With these data, the platform can give the communication features in different kind of people's needs. They can have different ways to use it. The registration page will create a certain account in the platform so the users can use the chat features there. Once it's registered successfully, users can login to the platform.

### 4.4.3 Dashboard

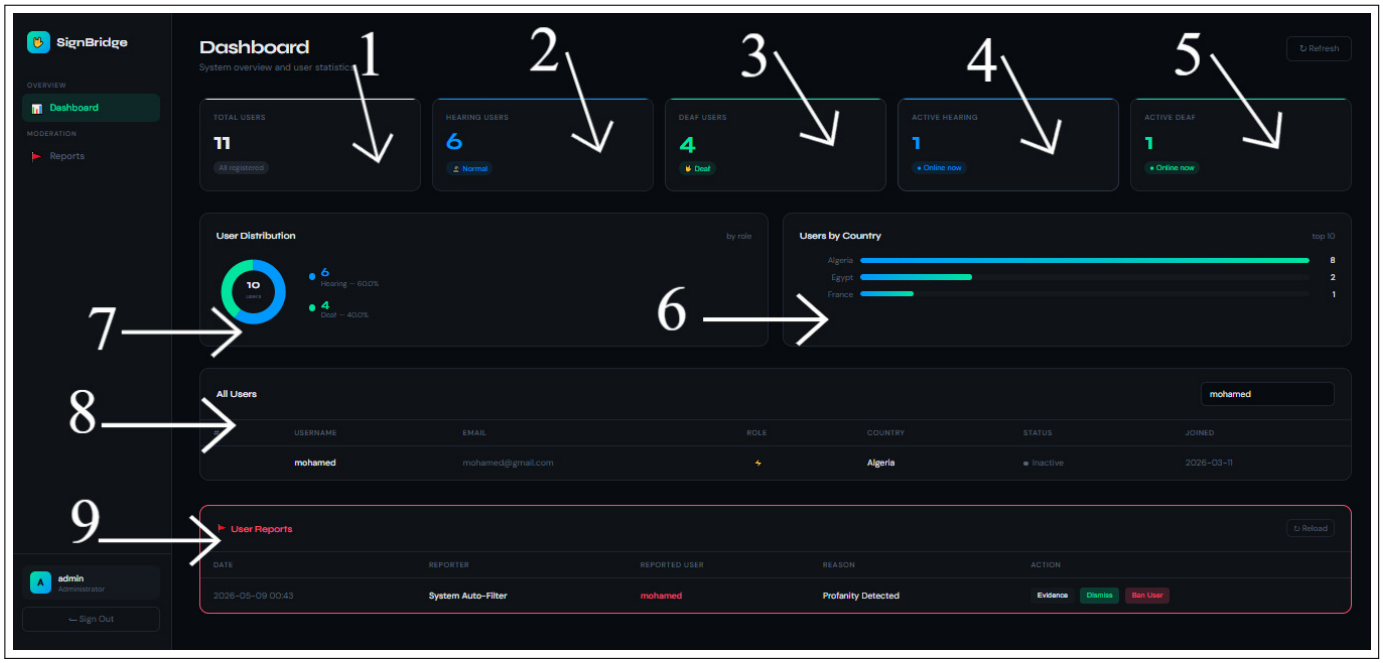



**Figure 4.5:Dashboard page view.**

Figure 4.5 shows us DashBoard of Admin. This dashboard include:

1. **Total users:** Total number of users who have been registered in the system (both hearing users and deaf users).
2. **Hearing users:**Number of hearing users who are registered in the system.
3. **Deaf users:**Total number of deaf users who are registered in the system.
4. **Active hearing users:** Number of hearing users who are currently in the system and using it.
5. **Active deaf users:** Number of deaf users who are currently in the system and using it.
6. **User distribution by country:**Shows the countries where users are from and represents the number of users in each country using a graph.
7. **User distribution by role:** Shows the distribution proportion of different types of users (including hearing and deaf users) on the platform using a circular diagram.
8. **All users table:** A table of all registered users' details like name, e-mail, role, country, and activation status, etc. The search box could help the admin to quickly search for the required users to manage accounts.
9. **User reports:** List of user reports including reporter, the reporter user and reported user, the reasons and the corresponding admin actions like reject or ban. It help the admin to check for misbehavior and to ban users to guarantee the safety of the platform.

### 4.4.4 User settings

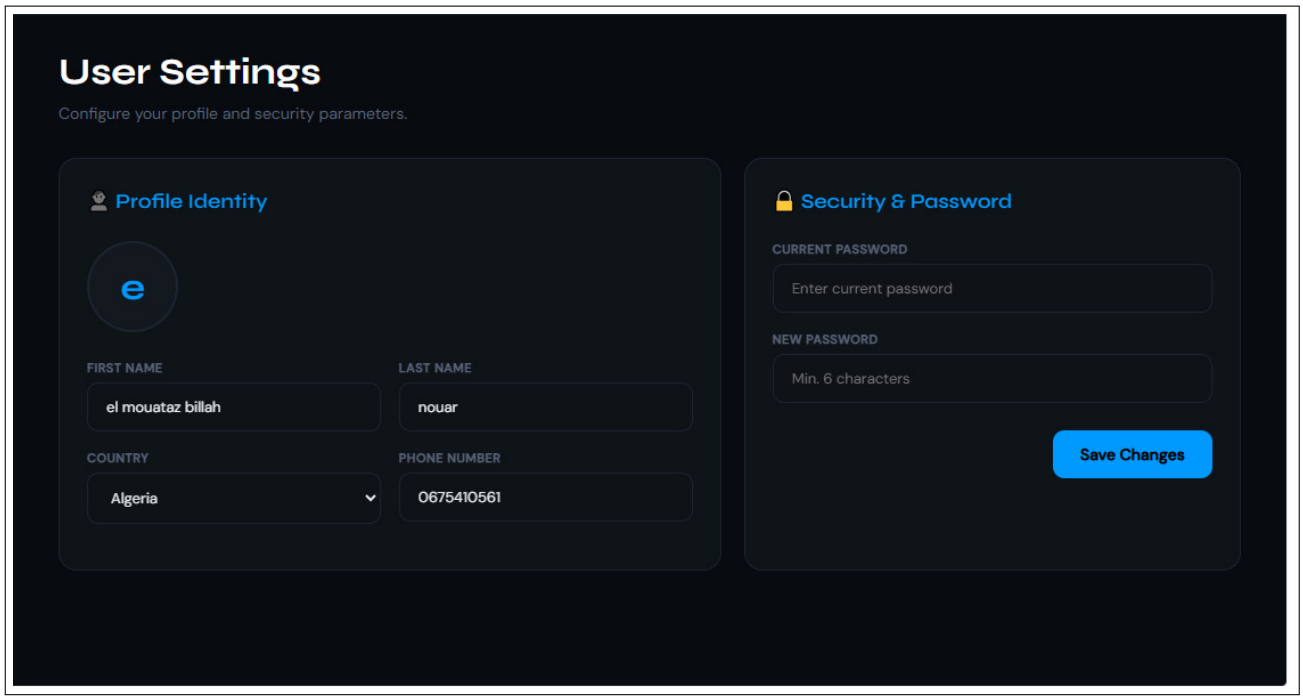



**Figure 4.6:User Settings page view.**

The settings page is an interface that gives users full control over their personal account details as well as security settings,as shown in the figure 4.6. Using this interface, users are given full control to edit any aspect of their details, i.e., first name, last name, country, phone number and their password at any point of time. This allows for accuracy and updates to user data and overall usability of the site to be improved.

### 4.4.5 Normal person chat




**Figure 4.7:Normal person chat page view.**

The normal chat interface was created to provide a fluid and real time chatting interface for the user as shown in figure 4.7. The page has been divided into 5 parts:

1. **Contact management:**The users can access the contact management through the search bar. The user can search, add or remove contacts and view the online or offline status of each contact present.

2. **Settings and Sign Out:**Users can navigate to other pages such as the settings page or sign out of the system. Through the settings page, they can update their personal information, while the logout button allows them to securely exit the system.

3. **Conversation area:**The conversations are displayed in the form of text messages with time stamp so that the users can have the actual feel of the real time conversation. The chat area displays the messages in a read and user-friendly format.

4. **Message input:**The user can enter their message into the input text box provided at the bottom of the page and hit enter to send the message to their friend with a guarantee of confidentiality in transmission.

5. **Reporting from Users:** It is a section where user can report the undesirable content or misbehave behavior occurred during conversations. There will be predefined report options such as inappropriate language, harassment, abuse, spamming, etc to select from, and then report the issue to administrator of the system. This function would improve the security of the conversations and the safety of the users , as shown in figure 4.8 .

   All reports that are being submitted by user will be saved into the admin's local json file instead of storing them in the relational database. When the report is being generated, it will auto save the previous five messages from the context of the conversations. So even if the reported user account was deleted, the report history remains still accessible from JSON file.

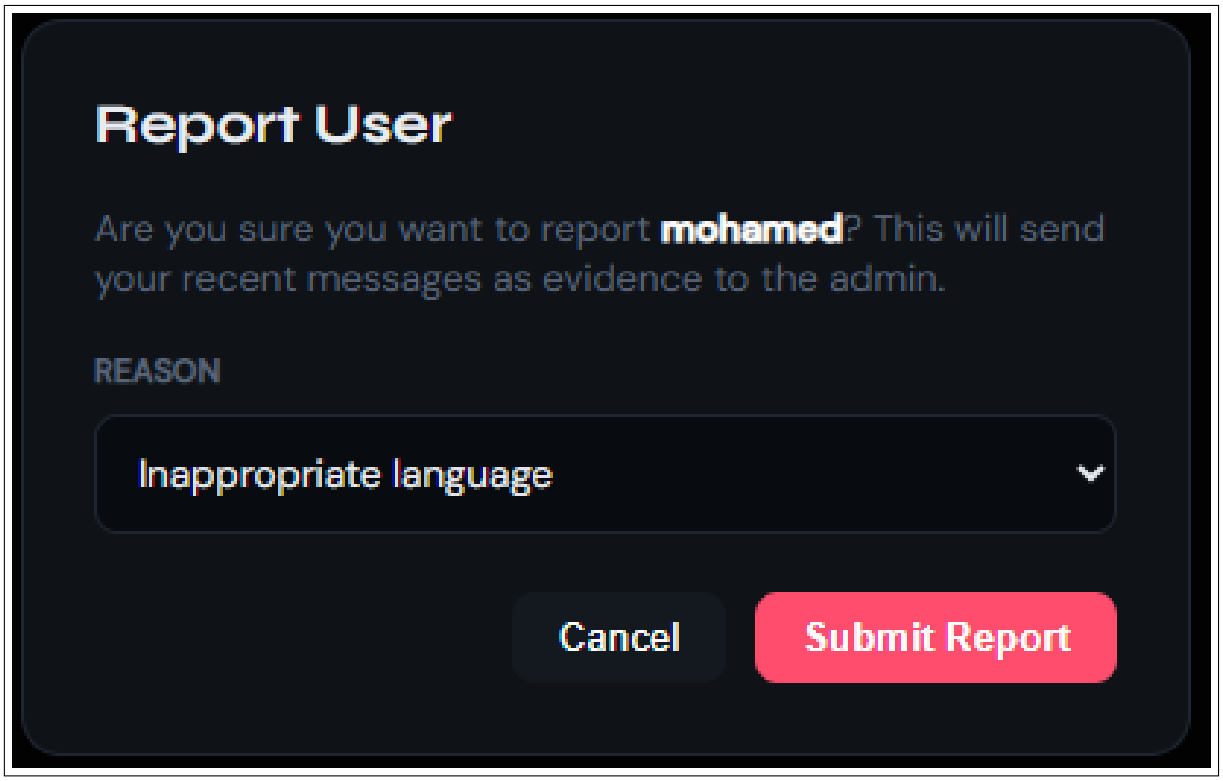



**Figure 4.8:User Reporting view.**

### 4.4.6 Deaf person chat

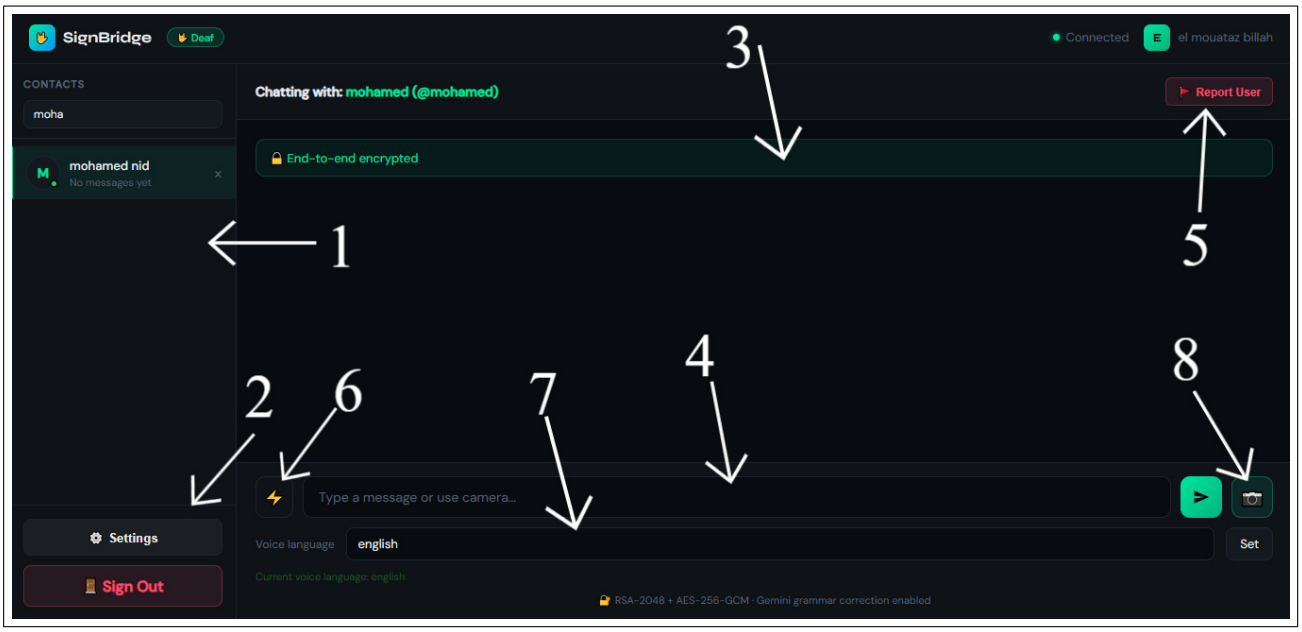



**Figure 4.9:Deaf person chat page view.**

This was a communicative and efficient chatting interface built specifically for deaf users to interact among themselves. This user interface consists of different functional sections that helps users interact, navigate and communicate in a real-time environment.

The first five features on the deaf communication interface are the same as on the normal chat interface: managing contacts, navigating the interface, displaying conversations, typing messages, and reporting functions. Therefore only the accessibility-specific features for deaf users are provided in this section, the Quick Communication feature, Voice Language Configuration section and Camera Activation and Sign Recognition section, as shown in figure 4.9:

6. **Quick communication:**This button gives quick access to predefined shorter messages like 'Yes' or 'No', emergency messages 'SOS' or 'OK' and greets with 'Hello', that aims to increase the speed of communication for users.
7. **Voice Language Configuration:**This part helps the users in changing the language of the voice output and speech enabled functions on the application, with default setting as English and a choice for users to choose and set the voice language to another one if needed.
8. **Camera activation and Sign Recognition:**This is section where camera is activated and used to perform sign language recognition in real time. Once the camera is turned ON,it will start detecting hand gestures at every moment and outputs will be given in real-time.

- system uses Top-1 prediction result and show the top 5 to identify each recognized sign within the camera space and will automatically write the result in the sentence section below the camera. Users can pause button to stop the detection and predictions, meaning that the user will not be given any new results in the sentence area while the pause button is clicked and can remove the wrongly recognized word Manually, as shown is figure 4.10:

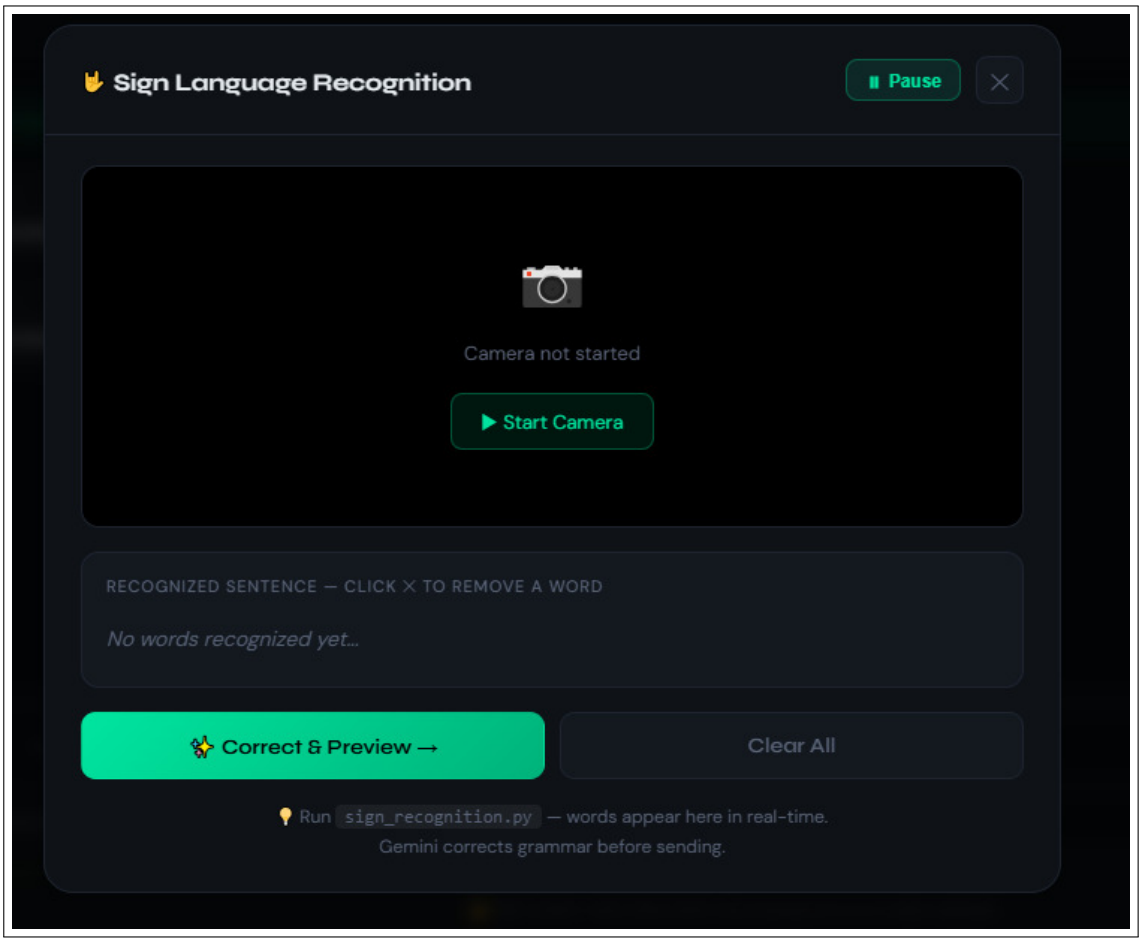



**Figure 4.10:Deaf person sign recognition view.**

## 4.5 Conclusion

The implementation and testing of the proposed system are detailed in this chapter. It covers the development platform, including hardware components, software, programming language, and libraries used. The effectiveness of the model is tested by considering aspects such as the training process, as well as the outcomes. Besides, the application user interface is described by highlighting some important features, including authentication, dashboard, user settings, and communication pages for both deaf users and normal users.

# General Conclusion

In this thesis, a web application is developed to enable effective and systematic interaction between deaf, non-speaking, and hearing people. The work is divided into four main chapters, each of which deals with an essential aspect of the proposed system.

The first chapter presented the basic concepts of sign language, including its definition, classification, types, and practical applications. It also highlighted the main challenges in communication between deaf and hearing people, and reviewed relevant sign language datasets and their uses in assistive technologies.

The second chapter presented the theoretical background of machine learning and deep learning methods applied in this field. He discusses modern artificial intelligence techniques, including supervised learning models, transformers, large language models, and computer vision systems, while reviewing the relevant scientific literature to define the context of the research.

The third chapter focused on system design and data processing. The selected dataset and processing techniques are described in detail. The overall system architecture and interaction workflow were defined, and the application architecture was modeled using UML diagrams, including use case, class, sequence, and activity diagrams. In addition, the database design was developed using entity relationship modeling to ensure a cohesive and efficient data structure.

Chapter Four presents the implementation and evaluation of the proposed system. The development environment comprises both hardware and software resources used during the implementation of the system. The software side is based on front-end technologies (HTML, CSS, and JavaScript), while the back-end is implemented using Python and Flask, along with the required libraries. This chapter also details the model training process, the evaluation methodology, and the obtained performance results, which are analyzed and discussed. In addition, the application interface was assessed through its main components, including authentication pages, the control panel, user settings, and communication modules.

Overall, this project represented an important step towards a better understanding of artificial intelligence, web development and assistive technologies. The final system achieved its main objective of bridging the communication gap between deaf and non-speaking people and those with normal hearing. To implement this project in the future, further efforts are still needed to improve the model's accuracy, expand the dataset, and reduce response times and evaluating the system across multiple languages.

# Bibliography


[1]Brentari, D. (2010). *Sign Languages: A Cambridge Language Survey.* Cambridge University Press. Google Books.

[2]Camgoz, N. C., Hadfield, S., & Bowden, R. (2018). *Neural Sign Language Translation.* IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[3]Othman, A. (2024). *Sign Language Processing: From Gesture to Meaning.* Springer, Cham.

[4]Way, A., Leeson, L., & Shterionov, D. (Eds.). (2024). *Sign Language Machine Translation.* Springer, Cham.

[5]Stokoe, W. C. (1960). *Sign Language Structure: An Outline of the Visual Communication System of the American Deaf.* Studies in Linguistics: Occasional Papers No. 8. Washington, D.C.: Gallaudet University.

[6]Klima, E. S., & Bellugi, U. (1979). *The Signs of Language.* Harvard University Press. Google Books.

[7]Baker-Shenk, C., & Cokely, D. (1980). *American Sign Language: A Teacher's Resource Text on Grammar and Culture.* Linstok Press. Google Books.

[8]Zeshan, U., & de Vos, C. (2012). *Sign Languages in Village Communities: Anthropological and Linguistic Insights.* De Gruyter Mouton. Google Books.

[9]de Vos, C. (2012). *Sign-Spatiality in Kata Kolok: How a Village Sign Language of Bali Inscribes Its Signing Space.* Nijmegen: Radboud University.

[10]Nyst, V. (2007). *A Descriptive Analysis of Adamorobe Sign Language (Ghana).* Utrecht: LOT Publications.

[11]Rosenstock, R., & Napier, J. (2016). *International Sign: Linguistic, Usage, and Status Issues.* Washington, DC: Gallaudet University Press.

[12]Deafblind International. (2022). *Deafblind Reality in Africa: New Research from DBI's African Researchers' Initiative (ARI).* Deafblind International.

[13]National Deaf Children's Society. (2025). *Tactile signing: Communication options for Deafblind children.* National Deaf Children's Society.

[14]Collins, S., & Petronio, K. (1998)."What Happens in Tactile ASL?" In C. Lucas (Ed.), "Pinky Extension and Eye Gaze: Language Use in Deaf Communities" (pp. 18–37). Washington, DC: Gallaudet University Press.

[15] Microsoft Research. (2023). *ASL Citizen – A Community-Sourced Dataset for Advancing Isolated Sign Language Recognition.*

[16] Khellas, K. (2023). *Alabib-65: Algerian Sign Language Dataset.* Mendeley Data.

[17] Luqman, H. (2023). *ArabSign: A Multi-modality Dataset and Benchmark for Continuous Arabic Sign Language Recognition.* IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG).

[18] Forster, J., Schmidt, C., Koller, O., Bellgardt, M., & Ney, H. (2014). *Extensions of the Sign Language Recognition and Translation Corpus RWTH-PHOENIX-Weather.* Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC).

[19]Zhou, H., Zhou, W., Qi, W., Pu, J., & Li, H. (2021). *CSL-Daily: Continuous Chinese Sign Language Dataset for Recognition and Translation.* Dataset homepage.

[20] Koller, O., Camgoz, N. C., García, I., & Ney, H. (2020). *How2Sign: A Large-Scale Multimodal Dataset for Continuous American Sign Language Translation.* In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

[21] Latif, G., Alghazo, J., Nazeeruddin, M., AlKhalaf, R., & AlKhalaf, R. (2018). *Arabic Alphabets Sign Language Dataset (ArASL).* Mendeley Data. DOI:10.17632/y7pckrw6z2.1

[22] Sarker, M. I.A., & collaborator(s). (2024). *Images of American Sign Language (ASL) Alphabet Gestures.* Mendeley Data. DOI:10.17632/48dg9vhmyk.1

[23] Wan, J., Zhao, Y., Zhou, S., Guyon, I., Escalera, S., & Li, S. Z. (2016). *ChaLearn LAP RGB-D Isolated and Continuous Gesture Datasets for Gesture Recognition.* CVPR Workshop.

[24] Luqman, H. and El-Alfy, E.-S. M. (2021). *Towards Hybrid Multimodal Manual and Non-Manual Arabic Sign Language Recognition: mArSL Database and Pilot Study.* Electronics, 10(14), 1739. DOI:10.3390/electronics10141739

[25] Hand Talk. (n.d.). *Hand Talk: Sign Language Translator and Learning App.* Hand Talk.

[26] Lingvano. (n.d.). *Lingvano: Online Sign Language Learning Platform.* Lingvano.

[27] ASL Dictionary. (n.d.). *Comprehensive American Sign Language Video Dictionary.* Software Studios.

[28] Sorenson ntouch. (n.d.). *Video Relay Service (VRS) Application for Deaf and Hard-of-Hearing Users.* Sorenson Communications.

[29] Rodríguez-Correa, P. A., Valencia-Arias, A., Patiño-Toro, O. N., Oblitas Díaz, Y., & Teodori De la Puente, R. (2023). *Benefits and development of assistive technologies for Deaf people's communication: A systematic review.* Frontiers in Education, 8, 1121597.

[30] Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*(4th ed.). Pearson.

[31] Zhang, W., Shen, X., Zhang, H., et al. (2024). *Feature importance measure of a multilayer perceptron based on the presingle-connection layer.* Knowledge and Information Systems, 66, 511–533.

[32] GeeksforGeeks. (n.d.). *Support Vector Machine Algorithm.* Accessed on March 26, 2026.

[33] Scikit-learn. (n.d.). *Support Vector Machines (SVM) Documentation.* Accessed on March 26, 2026.

[34] GeeksforGeeks. (n.d.). *Multi-Layer Perceptron Learning in TensorFlow.* Accessed on March 26, 2026.

[35] DataCamp. (n.d.). *Multilayer Perceptrons in Machine Learning.* Accessed on March 26, 2026.

[36] GeeksforGeeks. (n.d.). *Random Forest Algorithm in Machine Learning.* Accessed on March 26, 2026.

[37] GeeksforGeeks. (n.d.). *Decision Tree in Machine Learning.* Accessed on March 26, 2026.

[38] GeeksforGeeks. (n.d.). *Convolutional Neural Network (CNN) in Machine Learning.* Accessed on March 26, 2026.

[39] GeeksforGeeks. (n.d.). *Introduction to Recurrent Neural Networks (RNN).* Accessed on March 26, 2026.

[40] IBM. (n.d.). *Introduction to Long Short-Term Memory (LSTM) Networks.* Accessed on March 26, 2026.

[41] IBM. (n.d.). *Large Language Models Overview.* Accessed on March 26, 2026.

[42]Brown, T. B., Mann, B., Ryder, N., et al. (2020). *Language Models are Few-Shot Learners.*

[43] Team, G. D. (2023). *Gemini: A Family of Highly Capable Multimodal Models.*

[44] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*

[45] Raffel, C., Shazeer, N., Roberts, A., et al. (2020). *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer.* Journal of Machine Learning Research, 21(140), 1–67.

[46] Chowdhery, A., Narang, S., Devlin, J., et al. (2022). *PaLM: Scaling Language Modeling with Pathways.*

[47] Touvron, H., Lavril, T., Izacard, G., et al. (2023). *LLaMA: Open and Efficient Foundation Language Models.*

[48] Bai, Y., Kadavath, S., Kundu, S., et al. (2022). *Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback.*

[49] Jiang, A. Q., Sablayrolles, A., Mensch, A., et al. (2023). *Mistral 7B*

[50] GeeksforGeeks. (n.d.). *Getting Started with Transformers.* Accessed on March 28, 2026.

[51] IBM. (n.d.). *Introduction to Transformer Models.* Accessed on March 28, 2026.

[52] GeeksforGeeks. (n.d.). *Computer Vision Overview.* Accessed on March 28, 2026.

[53] GeeksforGeeks. (n.d.). *Main Steps in a Typical Computer Vision Pipeline.* Accessed on March 28, 2026.

[54] Bradski, G. (2000). *The OpenCV Library.* Dr. Dobb's Journal of Software Tools.

[55] Lugaresi, C., Tang, J., Nash, H., et al. (2019). *MediaPipe: A Framework for Building Perception Pipelines*

[56] Marcel, S., Rodriguez, Y., et al. (2010). *TorchVision: Datasets, Transforms and Models for Computer Vision.* PyTorch Documentation.

[57] TensorFlow Team. (2017). *TensorFlow Lite: Lightweight Machine Learning for Mobile and Embedded Devices.* Google.

[58] Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2019). *OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields.* IEEE TPAMI.

[59] Wu, Y., Kirillov, A., Massa, F., et al. (2019). *Detectron2.* Facebook AI Research (FAIR).

[60] Intel Corporation. (2018). *OpenVINO Toolkit: Open Visual Inference and Neural Network Optimization.* Intel.

[61] Hussin, S. K., Mohamed, O., Mohamed, M., Ahmed, E., & Mahmoud, O. (2024). *Real-time Arabic Sign Language Translator Using MediaPipe and LSTM.* PLOMS AI, 4(1).

[62] Kumari, D., & Anand, R. S. (2024). *Isolated Video-Based Sign Language Recognition Using a Hybrid CNN-LSTM with Attention Mechanism.* Electronics, 13(7), 1229.

[63] Utomo, W., Suhanda, Y., Ar-Rasyid, H., & Dharmalau, A. (2025). *Indonesian Language Sign Detection using Mediapipe with Long Short-Term Memory (LSTM) Algorithm.* Journal of Informatics and Web Engineering, 4(3), 245–258.

[64] Uddin, M. Z., Boletsis, C., & Rudshavn, P. (2025). *Real-Time Norwegian Sign Language Recognition Using MediaPipe and LSTM.* Multimodal Technologies and Interaction, 9(3), 23.

[65] Abdul Ameer, R. S., Ahmed, M. A., Al-Qaysi, Z. T., Salih, M. M., & Shuwandy, M. L. (2024). *Empowering Communication: A Deep Learning Framework for Arabic Sign Language Recognition with an Attention Mechanism.* Computers, 13(6), 153.

[66] Noor, T. H., Noor, A., Alharbi, A. F., Faisal, A., Alrashidi, R., Alsaedi, A. S., Alharbi, G., Alsanoosy, T., & Alsaeedi, A. (2024). *Real-Time Arabic Sign Language Recognition Using a Hybrid Deep Learning Model.* Sensors, 24(11), 3683.

[67] Jayanthi, P., Ponsy R. K. S., & Madhubalasri, B. (2023). *Sign Language Recognition using Deep CNN with Normalised Keyframe Extraction and LSTM.* Journal of Scientific & Industrial Research (JSIR).

[68] AI Studio. (n.d.). *API Keys Documentation.* Accessed on April 8, 2026.

[69] *Alex Nordeen .( 2020).Learn UML in 24 Hours.*

[70] Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual.* Scotts Valley, CA: CreateSpace.

[71] Duckett,J.(2011).*HTML and CSS: Design and Build Websites.*

[72] Meyer, E., & Weyl, E. (2017). *CSS: The Definitive Guide: Visual Presentation for the Web.*

[73] Codecademy. (n.d.). *What is JavaScript?* Accessed on April 8, 2026.

[74] Built In. (n.d.). *Introduction to MySQL.* Accessed on April 8, 2026.

[75] IONOS. (n.d.). *XAMPP Tutorial: Create Your Own Local Test Server.* Accessed on April 8, 2026.

[76] phpMyAdmin. (n.d.). *phpMyAdmin Official Website.* Accessed on April 8, 2026.

[77] PyTorch. (n.d.). *PyTorch Project Overview.* Accessed on April 8, 2026.

[78] MediaPipe. (n.d.). *MediaPipe Official Website.* Accessed on April 8, 2026.

[79] OpenCV. (n.d.). *About OpenCV.* Accessed on April 8, 2026.

[80] Kern-IT. (n.d.). *Flask Definition.* Accessed on April 8, 2026.

[81] GeeksforGeeks. (n.d.). *Introduction to MoviePy.* Accessed on April 8, 2026.

[82] Cryptography. (n.d.). *Python library for cryptographic recipes and primitives.*Accessed on April 8, 2026.

[83] bcryptjs. (n.d.). *JavaScript library for secure password hashing.*Accessed on April 8, 2026.

[84] GeeksforGeeks. (n.d.). *Socket Programming in Python.* Accessed on March 26, 2026.

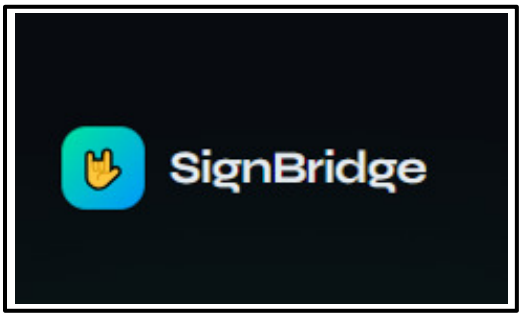
[85] GeeksforGeeks. (n.d.). *Loss Functions in Deep Learning.* Accessed on April 9, 2026.

[86] GeeksforGeeks. (n.d.). *Confusion Matrix in Machine Learning.* Accessed on April 9, 2026.

الجمهورية الجزائرية الديمقراطية الشعبية وزارة التعليم العالي والبحث العلمي
جامعة محمد خيضر بسكرة

# مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم والأشخاص غير القادرين على الكلام

مشروع لنيل شهادة مؤسسة ناشئة في اطار القرار الوزاري 1275




الاسم التجاري

**جسر الإشارة**

السنة الجامعية: 2026/2025

**مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم والأشخاص غير القادرين على الكلام**

## بطاقة معلومات:

## حول فريق الاشراف وفريق العمل

### 1- فريق الاشراف:

| فريق الاشراف | |
|---|---|
| **المشرف الرئيسي :**<br>**الأستاذة حميدي الزهرة**<br>**المشرف المساعد:**<br>**الأستاذة بن طراح أحلام** | **التخصص:**<br>**اعلام الي** |

### 2- فريق العمل:

| فريق المشروع | التخصص | الكلية |
|---|---|---|
| **الطالب:**<br>**نوار المعتز بالله** | اعلام الي تخصص<br>ذكاء اصطناعي | علوم الدقيقة والطبيعة والحياة |
| **الطالب:**<br>**نيد محمد القريشي** | اعلام الي تخصص<br>ذكاء اصطناعي | علوم الدقيقة والطبيعة والحياة |

# مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم والأشخاص غير القادرين على الكلام

## فهرس المحتويات


المقدمة: ..... 4
المحورالأول: تقديمالمشروع ..... 5
1. فكرة المشروع(الحل المقترح) : ..... 6
2. القيم المقترحة: ..... 7
3. فريق العمل: ..... 8
4. أهداف المشروع: ..... 8
5. الجدول الزمني لتحقيق المشروع ..... 9
المحورالثاني : الجوانب الابتكارية ..... 10
1. طبيعة الابتكارات ..... 11
2. مجالات الابتكار ..... 11
المحور الثالث: التحليل الاستراتيجي للسوق ..... 13
1. تحليل المتغيرات الكلية (PESTEL) : ..... 14
2. تحليل القوى التنافسية (PORTER) : ..... 15
3.تحليل (SWOT): ..... 15
4.عرض القطاع السوقي: ..... 16
5.قياس شدة المنافسة: ..... 17
6. الاستراتيجيات التسويقية: ..... 17
المحور الرابع: خطة الإنتاج والتنظيم ..... 19
1. عملية الانتاج: ..... 20
2. التموين: ..... 22
3. اليد العاملة: ..... 22
4. الشراكات الرئيسية: ..... 22
المحور الخامس: الخطةالمالية ..... 23
1.التكاليف والأعباء: ..... 24
2. رقم الاعمال: ..... 25
3. جدول حسابات النتائج المتوقع: ..... 25
4. خطة الخزينة: ..... 26
المحور السادسالنموذج الاولي التجريبي ..... 27

# مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم والأشخاص غير القادرين على الكلام

## المقدمة:

تعتبر المؤسسات الناشئة من أهم المشاريع التي تساهم في تجسيد الأفكار المبتكرة وتحويلها إلى حلول حقيقية تخدم المجتمع وتلبي احتياجات الأفراد. ومع التطور الكبير في مجال الذكاء الاصطناعي وتقنيات معالجة اللغة الطبيعية والرؤية الحاسوبية، ظهرت حلول حديثة تهدف إلى تسهيل التواصل وتحسين جودة الحياة لمختلف الفئات، خاصة فئة الصم والأشخاص غير القادرين على الكلام.

ورغم هذا التطور، ما تزال هذه الفئة تواجه صعوبات في التواصل اليومي مع الآخرين، سواء داخل المؤسسات أو في الحياة اليومية، مما يخلق حاجزا يؤثر على اندماجهم الاجتماعي وسهولة تفاعلهم مع المجتمع. كما أن الاعتماد على الطرق التقليدية أو مترجمي لغة الإشارة ليس متاحا دائما، وقد يتطلب وقتا وجهدا إضافيا.

لذلك قمنا بإنشاء منصة **ساين بريدج(Sign Bridge)** ، وهي تطبيق ويب ذكي يهدف إلى تسهيل التواصل بين الصم والأشخاص العاديين من خلال نظام محادثة تفاعلي يعتمد على تقنيات الذكاء الاصطناعي ونماذج التعرف على لغة الإشارة. حيث تقوم المنصة بتحليل إشارات المستخدم وترجمتها إلى نصوص مفهومة، إضافة إلى توفير وسائل تواصل مختلفة تساعد على جعل عملية التواصل أكثر سرعة وسلاسة وفعالية، مما يساهم في تحسين الاندماج الرقمي والاجتماعي لهذه الفئة.

# المحورالأول: تقديم المشروع

## مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم والأشخاص غير القادرين على الكلام

### 1. فكرة المشروع(الحل المقترح):

- مجال نشاطنا يتمثل في الخدمات الرقمية والذكاء الاصطناعي، من خلال تطوير منصة رقمية ذكية للتواصل بين الصم والأشخاص غير القادرين على الكلام مع الأشخاص العاديين.
- بدأت الفكرة من التحديات التي تواجه فئة الصم في التواصل اليومي داخل المؤسسات التعليمية والإدارية وفي الحياة العامة، حيث يعاني الكثير منهم من صعوبة إيصال أفكارهم أو فهم الآخرين بسبب غياب وسائل تواصل ذكية وسهلة الاستخدام. كما أن الاعتماد على مترجمي لغة الإشارة ليس متاحا دائما، مما يسبب بطئا وصعوبة في التفاعل. ومع التطور الكبير في تقنيات الذكاء الاصطناعي والرؤية الحاسوبية، جاءت فكرة إنشاء منصة تعتمد على نماذج التعرف على لغة الإشارة لتسهيل التواصل بشكل مباشر وفوري.
- يهدف مشروعنا إلى إنجاز منصة رقمية ذكية يتم من خلالها:

  - التعرف على لغة الإشارة: تحليل حركات اليد والإشارات باستخدام الذكاء الاصطناعي وتحويلها إلى نصوص مفهومة
  - المحادثة التفاعلية: توفير نظام محادثة بين المستخدمين يتيح التواصل بسهولة بين الصم والأشخاص العاديين.
  - تحويل النصوص والكلام: دعم تحويل النصوص أو الكلام إلى وسائل تواصل تساعد المستخدم الأصم على فهم الرسائل بشكل أسرع.
  - تسهيل الاندماج الاجتماعي: المساهمة في تقليل الحواجز التواصلية وتحسين التفاعل داخل المجتمع والمؤسسات.

- سيتم ذلك من خلال إنشاء تطبيق ويب يعتمد على تقنيات الذكاء الاصطناعي والرؤية الحاسوبية، والتعلم العميق، بالإضافة إلى استخدام نماذج لغوية كبيرة (LLM) لتحسين جودة التفاعل وفهم الرسائل بين المستخدمين.
- سيتم توجيه هذه المنصة للاستعمال داخل المؤسسات التعليمية، والإدارات العمومية، والمستشفيات، ومراكز الخدمات، إضافة إلى إمكانية استخدامها من طرف الأفراد في حياتهم اليومية لتسهيل التواصل مع فئة الصم والأشخاص غير القادرين على الكلام.

**2. القيم المقترحة:**

تتوافق القيم المقترحة أو المقدمة للعملاء في هذا المشروع مع العناصر التالية:

1) الحداثة: تلبية احتياجات فئة الصم والأشخاص غير القادرين على الكلام من خلال توفير منصة ذكية تعتمد على أحدث تقنيات الذكاء الاصطناعي، والرؤية الحاسوبية، والنماذج اللغوية الكبيرة(LLM) ، لتقديم حلول تواصل مبتكرة وحديثة.

2) الأداء: التأكد من أن أداء المنصة يلبي أو يتجاوز توقعات المستخدمين، من خلال توفير ترجمة سريعة ودقيقة للغة الإشارة، مع دعم المحادثات التفاعلية بشكل آني وآمن وموثوق.

3) المرونة: امتلاك القدرة على تطوير وتكييف النظام حسب احتياجات المستخدمين، مع إمكانية إضافة خصائص جديدة مستقبلا مثل دعم لغات إشارة متعددة أو تحسين دقة التعرف على الإشارات.

4) إنجاز المهام: مساعدة المستخدمين على إنجاز عمليات التواصل بسهولة، سواء داخل المؤسسات أو في الحياة اليومية، عبر تحويل لغة الإشارة إلى نصوص مفهومة وتسهيل التفاعل بين مختلف الأطراف.

5) التصميم: إنشاء واجهات استخدام حديثة وسهلة الاستعمال تتوافق مع احتياجات المستخدمين، مع التركيز على تجربة المستخدم وسهولة الوصول إلى مختلف خدمات المنصة.

6) خفض التكلفة: توفير حل رقمي يقلل الحاجة إلى مترجمي لغة الإشارة في بعض المواقف اليومية، مما يساهم في تقليل التكاليف وتوفير وسيلة تواصل متاحة لعدد أكبر من المستخدمين.

7) الحد من المخاطر: تطبيق تدابير أمنية لحماية بيانات المستخدمين وضمان خصوصية المحادثات والمعلومات الشخصية، مع توفير بيئة استخدام آمنة وموثوقة.

8) إمكانية الوصول :جعل خدمات المنصة متاحة بسهولة عبر تطبيق ويب يمكن استخدامه من مختلف الأجهزة، مما يسمح لفئة الصم والأشخاص غير القادرين على الكلام بالوصول إلى وسائل تواصل فعالة في أي وقت ومن أي مكان.

9) سهولة الاستخدام: تصميم النظام ليكون بسيطا وبديهيا، مما يساعد المستخدمين على التعامل معه دون الحاجة إلى خبرة تقنية كبيرة، ويزيد من رضاهم وسهولة تفاعلهم مع المنصة.

من خلال التركيز على هذه القيم، يهدف مشروع **ساين بريدج (Sign Bridge)** إلى تقديم نظام ذكي وحديث يساهم في تسهيل التواصل وتحسين الاندماج الرقمي والاجتماعي لفئة الصم والأشخاص غير القادرين على الكلام، مع توفير تجربة استخدام فعالة وآمنة تلبي احتياجات المستخدمين والمؤسسات المختلفة.

## 3. فريق العمل:

| الاسم و اللقب | المؤهلات و المهارات | المهام |
|---|---|---|
| **نوار المعتز بالله** | • ليسانس اعلام الي نظم معلوماتية<br>• 2 ماستر ذكاء اصطناعي | انشاء بنية تحتية قائمة على لغة الإشارة المختارة في تدريب النموذج.<br>تطوير الخادم الخلفي، إدارة قواعد البيانات، وربط مختلف خدمات النظام. |
| **نيد محمد القريشي** | • ليسانس اعلام الي نظم معلوماتية<br>• 2 ماستر ذكاء اصطناعي | تصميم وتطوير واجهات الاستخدام وتحسين تجربة المستخدم داخل المنصة.<br>تطوير وتدريب نماذج التعرف على لغة الإشارة وتحسين دقة النظام.<br>البحث والمعالجة المسبقة للبيانات لغة الإشارة. |

## 4. أهداف المشروع:

❖ تشمل الأهداف التجارية للمشروع ما يلي:

- **زيادة الوعي والتبني :**نشر الوعي حول أهمية تقنيات الذكاء الاصطناعي في تسهيل التواصل مع فئة الصم والأشخاص غير القادرين على الكلام، وتشجيع المؤسسات والأفراد على اعتماد المنصة في حياتهم اليومية.
- **تحسين التواصل :**توفير وسيلة تواصل ذكية وسريعة تساعد على تقليل الحواجز التواصلية بين الصم والأشخاص العاديين داخل المؤسسات التعليمية والإدارية والصحية وفي مختلف المجالات.
- **رضا المستخدمين :**ضمان مستويات عالية من رضا المستخدمين من خلال تقديم منصة موثوقة، سهلة الاستخدام، وسريعة الاستجابة ، تلبي احتياجات فئة الصم والأشخاص غير القادرين على الكلام.
- **قابلية التوسع :**إنشاء نظام قابل للتطوير يسمح بإضافة خصائص مستقبلية مثل دعم لغات إشارة متعددة، وتحسين دقة التعرف على الإشارات، والتكامل مع تطبيقات وخدمات أخرى.
- **الريادة :**نسعى إلى أن تصبح منصة ساين بريدج(Sign Bridge) من بين أبرز المنصات الذكية المتخصصة في تسهيل التواصل باستخدام لغة الإشارة على المستوى المحلي والدولي خلال السنوات القادمة.

## مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم والأشخاص غير القادرين على الكلام

- **الابتكار التقني :**تطوير منصة تعتمد على أحدث تقنيات الذكاء الاصطناعي، والرؤية الحاسوبية، والنماذج اللغوية الكبيرة (LLM) مع الاعتماد على بيانات لغة الإشارة الأمريكية (ASL Dataset) لتطوير نموذج دقيق وفعال للتعرف على الإشارات و لتقديم تجربة تواصل حديثة وفعالة.

**.الحصة السوقية المستهدفة:**

من خلال الاعتماد على تقنيات الذكاء الاصطناعي ونماذج التعرف على لغة الإشارة الأمريكية(ASL Dataset) ، يستهدف مشروع **ساين بريدج (Sign Bridge)** شريحة واسعة من المستخدمين داخل وخارج الجزائر، خاصة المؤسسات التعليمية، الصحية، والإدارية ، إضافة إلى الأفراد المهتمين بالتكنولوجيا المساعدة.

من المتوقع أن تحقق المنصة في مراحلها الأولى حصة سوقية تتراوح بين **10%و15 %**من المستخدمين المهتمين بحلول التواصل الذكية الخاصة بفئة الصم داخل السوق المحلية، مع إمكانية نمو تدريجي على المدى المتوسط نتيجة زيادة الوعي بالتكنولوجيا واعتماد المؤسسات للحلول الرقمية الحديثة.

## 5. الجدول الزمني لتحقيق المشروع

| المرحلة | النشاط | المدة الزمنية |
|---|---|---|
| 1. التخطيط والتحليل | دراسة فكرة المشروع وتحديد أهداف النظام | أسبوعان |
| | البحث في الدراسات السابقة وتحليل الأنظمة المشابهة | أسبوعان |
| | البحث عن مجموعات البيانات (Datasets) المناسبة | شهر |
| | معالجة البيانات وتجهيزها (Preprocessing) | أسبوعان |
| 2. التصميم والبرمجة | تصميم واجهات المستخدم وتجربة الاستخدام | 3 أسابيع |
| | تصميم معمارية النظام وربط المكونات | أسبوعان |
| | تطوير واجهة الويب (HTML, CSS, JavaScript) | شهر |
| | تطوير الخادم الخلفي باستخدام Flask | شهر |
| 3. تطوير الذكاء الاصطناعي والأنظمة الذكية | تدريب نموذج التعرف على لغة الإشارة | شهر |
| | دمج نموذج الـ LLM داخل النظام | أسبوع |
| | تطوير نظام التواصل الفوري بين المستخدمين | 3 أسابيع |
| | نظام توليد فيديو لغة الإشارة من الكلام (Generated Sign Language Video) | 3 أسابيع |
| 4. الاختبار والتحسين | اختبار النظام واكتشاف الأخطاء | أسبوعان |
| | تحسين الأداء ودقة النماذج | أسبوع |
| | اختبار تجربة المستخدم وإجراء التعديلات النهائية | أسبوع |
| 5. التوثيق والعرض | التوثيق (كتابة التقارير، شرح المراحل، توثيق الكود، دليل المستخدم) | مستمر |
| | تحضير العرض التقديمي وتحديثة بشكل مستمر | مستمر |
| 6. المتابعة والتطوير | جمع الملاحظات وتحليل الأداء | مستمر |
| | إضافة ميزات وتحسينات مستقبلية | مستمر |

# المحورالثاني : الجوانب الابتكارية

**مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم والأشخاص غير القادرين على الكلام**

## 1. طبيعة الابتكارات

يمكن تحديد طبيعة الابتكارات المعتمدة في المشروع على النحو التالي:

- **ابتكارات السوق**: تمثل في تطوير منصة رقمية ذكية موجهة لفئة الصم والأشخاص غير القادرين على الكلام، تهدف إلى تحسين أساليب التواصل التقليدية من خلال توفير حل حديث وسهل الاستخدام يعتمد على الذكاء الاصطناعي، مما يضمن سرعة التفاعل وسهولة التواصل مقارنة بالطرق التقليدية.

- **الابتكارات الجذرية**: يمثل المشروع تحولا جذريا في طريقة التواصل، من خلال اعتماد تقنيات التعرف على لغة الإشارة باستخدام نماذج الذكاء الاصطناعي والرؤية الحاسوبية، مما يسمح بتحويل الإشارات اليدوية إلى نصوص مفهومة بشكل فوري ودقيق،و انشاء فيديوات لغة إشارة من النصوص وبالتالي إلغاء الحاجة إلى وسيط بشري في أغلب الحالات.

- **الابتكارات التكنولوجية:**

- يعتمد المشروع على تقنيات متقدمة مثل التعلم العميق(Deep Learning) ، الرؤية الحاسوبية (Computer Vision)، والنماذج اللغوية الكبيرة(LLM) ، والتي تمكن من تحقيق دقة عالية في التعرف على لغة الإشارة وتحسين جودة التفاعل داخل المنصة بشكل آني وفعال.

- **الابتكارات المتزايدة**: يتضمن المشروع تحسين مستمر لنظام التعرف على الإشارات وواجهة المستخدم بناءً على ملاحظات المستخدمين، إضافة إلى تطوير دقة النموذج وتوسيع قاعدة البيانات(ASL Dataset) ، مع إمكانية إضافة ميزات جديدة مثل دعم لغات إشارة أخرى، وتحسين الأداء وتجربة المستخدم بشكل مستمر لضمان أفضل خدمة ممكنة.

## 2. مجالات الابتكار

- التعرف الذكي على لغة الإشارة

  - استخدام تقنيات الذكاء الاصطناعي والتعلم العميق لتحليل حركات اليد وتحويلها إلى نص مفهوم بشكل فوري.

  - تمكين التواصل المباشر بين مستخدمي لغة الإشارة والأشخاص العاديين دون الحاجة إلى مترجم بشري.

**مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم والأشخاص غير القادرين على الكلام**

- التواصل ثنائي الاتجاه
  - توفير منصة تسمح بتحويل لغة الإشارة إلى نص، وتحويل الكلام أو النص إلى فيديو بلغة الإشارة.
  - تسهيل التفاعل بين الصم وغير الناطقين مع مختلف فئات المجتمع.
- الأمان
  - حماية البيانات الشخصية والمحادثات من خلال تخزين آمن للمعلومات.
  - ضمان خصوصية المستخدمين أثناء استخدام الكاميرا والمحادثات داخل المنصة.
- الكفاءة
  - معالجة فورية للإشارات عبر الكاميرا في الزمن الحقيقي، مما يضمن سرعة التواصل.
  - تقليل الوقت المطلوب لفهم الرسائل بين الطرفين وتحسين تجربة الاستخدام.
- سهولة الاستخدام
  - تصميم واجهة بسيطة وسهلة الوصول لجميع الفئات، خاصة ذوي الاحتياجات الخاصة.
  - إمكانية استخدام المنصة مباشرة من المتصفح دون الحاجة إلى أجهزة إضافية.
- توفير التكاليف
  - تقليل الحاجة إلى مترجمي لغة الإشارة في المواقف اليومية البسيطة.
  - إتاحة حل رقمي منخفض التكلفة مقارنة بالخدمات التقليدية.
- الابتكار التقني
  - دمج الرؤية الحاسوبية مع نماذج اللغة الكبيرة (LLM) لتحسين فهم السياق وتصحيح الجمل الناتجة.
  - تطوير نظام توليد فيديو للغة الإشارة من النص أو الكلام، مما يمثل حلاً حديثًا غير متوفر بشكل واسع محليًا.

# المحور الثالث: التحليل الاستراتيجي للسوق

## 1. تحليل المتغيرات الكلية (PESTEL) :

**هي أداة تستخدم من طرف المؤسسات لتحليل ومراقبة العوامل البيئية التسويقية الخارجية التي تؤثر على هذه المؤسسات، ويتكون من العناصر التالية :**

| العوامل السياسية | العوامل الاقتصادية - | العوامل الاجتماعية |
|---|---|---|
| - سياسة دعم المشاريع الناشئة<br>. - السياسة الضريبية .<br>- سياسة تطوير المؤسسات.<br>- سياسة دعم الدولة<br>-التوجه نحو رقمنة الخدمات وتحسين الإدماج الاجتماعي لذوي الاحتياجات الخاصة.. | - تزايد الاستثمار في الحلول الرقمية والذكاء الاصطناعي.<br>-انخفاض تكاليف إطلاق المنصات الرقمية مقارنة بالمشاريع التقليدية.<br>-الحاجة المتزايدة لخدمات تسهل التواصل وتوفر الوقت والجهد. | -ارتفاع الوعي بأهمية دمج الصم وغير الناطقين في المجتمع.<br>- زيادة استخدام الإنترنت والتطبيقات الرقمية في الحياة اليومية.<br>- قبول المجتمع المتزايد للحلول التقنية المساعدة. |
| **العوامل التكنولوجية** | **العوامل البيئية** | **العوامل القانونية. -** |
| - تطور تقنيات الذكاء الاصطناعي والرؤية الحاسوبية بشكل سريع.<br>-توفر مكتبات وأدوات مفتوحة المصدر لتطوير أنظمة التعرف على لغة الإشارة. | -المشروع رقمي بالكامل، مما يقلل الحاجة إلى الموارد المادية.<br>-إمكانية العمل عن بعد دون أثر بيئي كبير. | -ضرورة احترام قوانين حماية البيانات الشخصية للمستخدمين.<br>-الالتزام بحقوق استخدام البيانات ومجموعات التدريب الخاصة بالنموذج.<br>-مراعاة قوانين الوصول الرقمي وإتاحة الخدمات لذوي الاحتياجات الخاصة.<br>-الالتزام بسياسات أمن المعلومات عند تخزين المحادثات والوسائط. |

.

## 2. تحليل القوى التنافسية (PORTER):

إطار بورتر يساعد على تقييم وضع المشروع في السوق من خلال دراسة خمس قوى تؤثر على المنافسة وجاذبية القطاع، وهو مناسب جدًا لمشروع Sign Bridge كمنصة رقمية موجهة للصم وغير الناطقين:

1. حدة المنافسة: متوسطة، لأنه لا توجد منصات جزائرية متخصصة بشكل مباشر في الترجمة الثنائية بين لغة الإشارة والنص/الكلام، لكن توجد بعض التطبيقات العالمية أو أدوات ترجمة عامة يمكن اعتبارها حلولًا جزئية.
2. قوة العملاء: متوسطة إلى مرتفعة، لأن المستخدمين لديهم توقعات عالية من حيث دقة الترجمة وسهولة الاستخدام، كما أن رضاهم يؤثر مباشرة على انتشار المنصة وقبولها.
3. قوة الموردين: متوسطة، لأن المشروع يعتمد على مصادر مثل مجموعات البيانات الخاصة بلغة الإشارة، خدمات الاستضافة السحابية، ومكتبات الذكاء الاصطناعي مفتوحة المصدر، وهي متوفرة لكن جودتها تختلف.
4. تهديد الخدمات البديلة: متوسط، إذ توجد بدائل مثل مترجمي لغة الإشارة البشريين أو تطبيقات المراسلة النصية، لكن المنصة تتميز بالترجمة الفورية والتفاعل المباشر المعتمد على الذكاء الاصطناعي.
5. تهديد دخول منافسين جدد: متوسط، لأن تطوير منصة مشابهة يتطلب خبرة في الذكاء الاصطناعي والرؤية الحاسوبية ومعالجة اللغة، إضافة إلى توفر بيانات تدريب مناسبة، مما يشكل حاجزًا تقنيًا أمام المنافسين الجدد.

## 3.تحليل (SWOT):

وهي طريقة للتحليل االستراتيجي تساعد على تحليل البيئة الداخلية للمؤسسة والمتمثلة بنقاط قوتها ونقاط ضعفها من جانب، وما يقابلها من تحليل للبيئة الخارجية المتمثلة بالفرص المتاحة والتهديدات التي قد تواجهها و قد سميت بهدا اختصرا لحرف الأول من كل عنصر من عناصرها .

| نقاط الضعف (Weakness) | نقاط القوة (Strengths) |
|---|---|
| • محدودية الموارد المالية كشركة ناشئة يمكن أن تحد من التسويق والتوسع.<br>• محدودية الوعي بالمنصة بسبب كونها خدمة جديدة<br>• لا تزال المنصة في مرحلة التطوير ولم يتم استعاملها على نطاق واسع.<br>• الحاجة إلى بيانات تدريب كبيرة ودقيقة لتحسين النموذج.<br>•اعتماد بعض الميزات على اتصال إنترنت جيد. | • خدمة فريدة توفر الامن والشفافية والسرعة في التواصل بين الطرفين<br>• تصميم مبتكر سهل الاستخدام ويمكن استعمالها في أي وقت .<br>• تساعد من تخفيض التكاليف و الوقت والسرعة.<br>• التكلفة المنخفضة للخدمة مقارنة مع الأعباء |

**مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم والأشخاص غير القادرين على الكلام**

| التهديدات (Threats) | الفرص (Opportunities) |
|---|---|
| • دخول شركات أو منصات عالمية تقدم خدمات مشابهة.<br>•صعوبة الحصول على بيانات محلية كافية للغة الإشارة الجزائرية.<br>•التقدم التكنولوجي الذي يسمح بظهور تكنولوجيات أخرى .<br>• مخاوف تتعلق بحماية البيانات والخصوصية.<br>• احتمال ضعف تبني التكنولوجيا لدى بعض الفئات في البداية. | • تزايد الاهتمام بالتكنولوجيا المساعدة لذوي الاحتياجات الخاصة.<br>• اهتمام المؤسسات الحكومية بالرقمنة<br>•التوسع نحو أسواق أخرى ودعم لغات إشارة متعددة..<br>• مكانية تطوير تطبيقات مرافقة للهاتف مستقبلاً. |

## 4.عرض القطاع السوقي:

❖ السوق المحتمل

✓ المستفيدون من خدماتنا

الأفراد من فئة الصم وغير الناطقين، عائلاتهم، المؤسسات التعليمية، المراكز الصحية، الإدارات العمومية، الجمعيات، والمؤسسات الخاصة التي تحتاج إلى تحسين التواصل مع هذه الفئة داخل وخارج الجزائر.

✓ دوافع الاستخدام

الحاجة إلى تسهيل التواصل اليومي، تقليل الحواجز بين الصم وغير الناطقين وباقي المجتمع، تحسين الوصول إلى الخدمات الأساسية (تعليم، صحة، إدارة)، والاستفادة من حل ذكي سريع يعتمد على الذكاء الاصطناعي والتواصل الفوري.

✓ أماكن التواجد

يتواجد المستخدمون في مختلف أنحاء العالم، داخل الجزائر وخارجها، خاصة في المدن التي تنتشر فيها الخدمات الرقمية، المؤسسات التعليمية والصحية، والجمعيات المهتمة بالإدماج الاجتماعي لذوي الاحتياجات الخاصة.

✓ حجم السوق

يشمل السوق شريحة كبيرة من الصم وغير الناطقين عالميًا، بالإضافة إلى عائلاتهم والمؤسسات المرتبطة بهم، مما يمثل سوقًا واسعًا وقابلًا للنمو مع تزايد الطلب على حلول التواصل المساعدة والتقنيات الذكية.

**مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم**
**والأشخاص غير القادرين على الكلام**

❖ الفئة المستهدفة

- ✓ **الأفراد من الصم وغير الناطقين :**نركز على توفير خدماتنا لهذه الفئة بهدف تسهيل التواصل اليومي، الدراسة، والعمل، والوصول إلى الخدمات العامة.
- ✓ **العائلات والأشخاص المحيطون بهم :**نستهدف الأشخاص الذين يتعاملون مع الصم وغير الناطقين لتسهيل التفاعل وفهم لغة الإشارة.
- ✓ **المؤسسات التعليمية والصحية :**نهدف إلى تزويد المدارس، الجامعات، والمراكز الصحية بحل يساعد على تحسين التواصل مع المستفيدين.
- ✓ **الإدارات والمؤسسات العمومية :**نسعى لتوفير وسيلة تساعد هذه الجهات على تقديم خدمات أكثر شمولية وسهولة لجميع المواطنين.

نستهدف هذه الفئات لأن مشروع **Sign Bridge** يهدف إلى تحسين جودة التواصل ورفع مستوى الإدماج الاجتماعي للصم وغير الناطقين. كما نركز على توسيع قاعدة المستخدمين من خلال الشراكة مع المؤسسات التعليمية، الصحية، والإدارية، مما يتيح للمشروع فرص نمو كبيرة وتحقيق قيمة اجتماعية واقتصادية في السوق المستهدف.

## 5.قياس شدة المنافسة:

تُعتبر شدة المنافسة في سوق تطبيقات التواصل الموجهة لذوي الاحتياجات الخاصة متوسطة إلى مرتفعة عالميًا، لكنها منخفضة محليًا في الجزائر بسبب محدودية الحلول المتخصصة. يوجد عدد من المنافسين المهمين مثل SignAllو Avaو Google Live Transcribeو Microsoft Translatorو Hand Talk، حيث تقدم هذه المنصات حلولًا تعتمد على الذكاء الاصطناعي لتحويل الكلام إلى نص أو ترجمة جزئية للغة الإشارة، وغالبًا ما تركز على جانب واحد من عملية التواصل وليس على نظام ثنائي متكامل.

- نقاط القوة لدى المنافسين:
  - ✓ استخدام تقنيات متقدمة مثل الذكاء الاصطناعي والتعرف على الصوت.
  - ✓ انتشار عالمي واسع وسهولة الوصول للمستخدمين.
  - ✓ موثوقية عالية خاصة لدى الشركات الكبرى مثل: Google.
- نقاط الضعف لدى المنافسين:
  - ✓ ضعف أو غياب دعم ثنائي الاتجاه الحقيقي للغة الإشارة.
  - ✓ عدم التخصص في تلبية احتياجات الصم بشكل كامل.
  - ✓ بعض الحلول تعتمد على خدمات مدفوعة أو تحتاج إنترنت قوي.

### 6. الاستراتيجيات التسويقية:

❖ **استراتيجية التوعية**

- o الاعتماد على التسويق المجاني عبر وسائل التواصل الاجتماعي.
- o نشر فيديوهات قصيرة توضح فكرة المشروع وطريقة استخدامه لجذب الانتباه.
- o المشاركة في مسابقات جامعية وحاضنات أعمال لزيادة شهرة المشروع بدون تكاليف كبيرة.

**مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم والأشخاص غير القادرين على الكلام**

- **استراتيجية التسويق منخفض التكلفة**
  - إنشاء صفحة رسمية للمشروع ونشر محتوى تعليمي وتوضيحي بشكل منتظم.
  - استخدام SEO بشكل أساسي لزيادة ظهور المشروع في البحث بدون إعلانات مدفوعة.

- **استراتيجية التسعير**
  - تقديم المشروع بشكل مجاني في المرحلة الأولى (لجذب المستخدمين).
  - لاحقًا يمكن اعتماد نموذج اشتراك منخفض التكلفة للمؤسسات فقط.

# المحور الرابع: خطة الإنتاج والتنظيم

**مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم والأشخاص غير القادرين على الكلام**

1. **عملية الانتاج:**

❖ اقتناء المواد الأولية
- ✓ جمع قواعد بيانات لغة الإشارة.
- ✓ توفير الصور والفيديوهات التدريبية.
- ✓ تجهيز الحواسيب والأدوات.
- ✓ تثبيت أدوات الذكاء الاصطناعي والرؤية الحاسوبية.
- ✓ تطوير بيئة العمل البرمجية.

❖ المعالجة المسبقة للبيانات:
- ✓ تنظيف البيانات وإزالة العينات غير الصالحة.
- ✓ تقسيم الفيديوهات إلى إطارات لاستخراج نقاط اليد والجسم.
- ✓ تطبيق التطبيع والتصنيف.

❖ تدريب نموذج الذكاء الاصطناعي:
- ✓ تدريب النموذج على التعرف على الإشارات.
- ✓ تحسين الأداء والدقة.

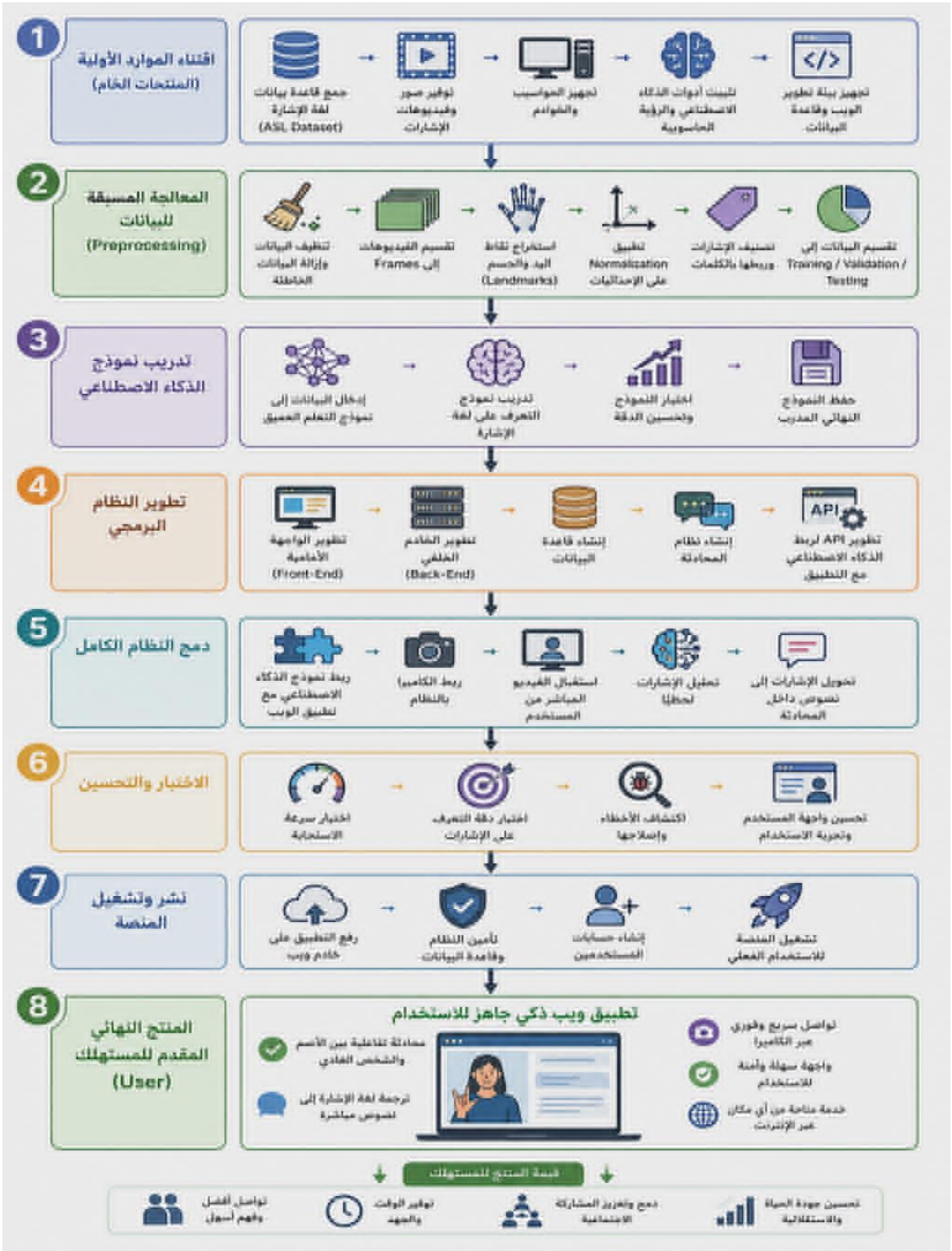
❖ تطوير النظام البرمجي:
- ✓ تطوير الواجهة الأمامية والخادم الخلفي.
- ✓ إنشاء قاعدة البيانات.
- ✓ إعداد نظام المحادثات.

❖ الاختبار والتحسين:
- ✓ اختبار سرعة الاستجابة.
- ✓ قياس دقة النموذج.
- ✓ اكتشاف الأخطاء.
- ✓ تحسين واجهة المستخدم.

❖ نشر وتشغيل المنصة:
- ✓ رفع التطبيق على الخادم.
- ✓ تأمين النظام.
- ✓ تسجيل المستخدمين.
- ✓ التشغيل النهائي.

- **مخطط مراحل الانتاج**

# مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم والأشخاص غير القادرين على الكلام



## 2. التمويل:

### ❖ سياسة الشراء:

- ✓ الاعتماد على الموارد المجانية ومفتوحة المصدر قدر الإمكان.
- ✓ شراء الخدمات الأساسية فقط مثل الاستضافة السحابية والنطاق عند الحاجة.
- ✓ اقتناء المعدات تدريجيًا حسب تطور المشروع واحتياجاته الفعلية.
- ✓ تقليل المصاريف بالاعتماد على العمل الجامعي والموارد المتاحة.

### ❖ أهم الموردين:

- ✓ مزودو الخدمات السحابية مثل Google Cloud و Amazon Web Services.
- ✓ منصات البيانات المفتوحة الخاصة بلغة الإشارة مثل Kaggle .
- ✓ مكتبات التطوير مثل TensorFlow و PyTorch و OpenCV.

### ❖ سياسة الدفع:

- ✓ في البداية الاعتماد على الخدمات المجانية أو التجريبية.
- ✓ الدفع الشهري للخدمات السحابية عند إطلاق المنصة رسميًا.
- ✓ شراء المعدات عند الحاجة وبالتدريج حسب الموارد المالية المتاحة.

### ❖ سياسة الاستلام:

- ✓ ستلام الخدمات الرقمية مباشرة عبر الإنترنت بعد الاشتراك.

## 3. اليد العاملة:

- ✓ مطورين ويبFull Stack .
- ✓ مختصين في التعلم العميق والرؤية الحاسوبية.
- ✓ مبرمجين متخصصين في التصميم والتنسيق.
- ✓ مستشارين لغة الإشارة.
- ✓ مسؤولين إدارة قاعدة البيانات.

## 4. الشراكات الرئيسية:

تتمثل الشراكات الرئيسية للمشروع في حاضنة الأعمال الجامعية التي تدعم التأسيس والتوجيه والتمويل وجمعيات الصم والبكم التي تساهم في الاختبار وتحسين التجربة وربط المشروع بالمستخدمين، ومديرية النشاط الاجتماعي والتضامن التي تدعم الإدماج والتوسع المؤسسي. كما يشمل المشروع شراكات مع شركات تقنية مثل Google و Microsoft و Metaلتوفير حلول الذكاء الاصطناعي والخدمات السحابية إضافة إلى مؤسسات تعليمية وبحثية تساهم في تطوير النموذج وتحسين أدائه علمياً وتقنياً.

# المحور الخامس: الخطة المالية

## 1.التكاليف والأعباء:

أولا- تكاليف المشروع واهتلاك الاستثمار :

- تكاليف المشروع: تتمثل التكاليف الاجمالية للمشروع في التكاليف الاستثمارية والتكاليف التشغيلية: التكاليف الاستثمارية:

| الأصول | التكلفة |
|---|---|
| الآلات والمعدات | -حواسيب شخصية (3أجهزة)450,000دينار جزائري<br>-حاسوب معالجة (تجربة ميدانية صغيرة 600,000 - 250,000 دج)<br>-التخزين والنسخ الاحتياطي (تجربة ميدانية صغيرة: 50,000 - 150,000 دج)<br>-الشبكة والكابلات(تجربة ميدانية صغيرة: 40,000 - 120,000 دج)<br>-الكاميرا أو كاميرتان (تجربة ميدانية صغيرة: 80,000 - 250,000 دج) |
| رأس المال العامل | - / |
| المجموع | **1,570,000** |

- الهيكل التمويلي: يتم تمويل المشروع بعدة طرق إما بالاعتماد الكلي على الأموال الخاصة لصاحب المشروع وهذا ما يسمى بالتمويل الذاتي، أو الاستعانة بأحد المؤسسات المالية وذلك عن طريق التمويل الثنائي او الثلاثي، أما بالنسبة لمشروعنا (نظام مساعد لضعاف البصر) سيكون بالتمويل الثنائي كما هو موضح في الجدول التالي:

**التكلفة التشغيلية** فتشمل المصاريف المستمرة مثل الصيانة، التحديثات، الدعم التقني، واستهلاك الإنترنت والكهرباء، ويمكن تقديرها بحوالي **150,000إلى 300,000 دج سنوياً**.

**التكلفة الإجمالية =** التكلفة الاستثمارية+ التكلفة التشغيلية **التكلفة الإجمالية =1,570,000+300,000= 1,870,000**

## 2. رقم الاعمال:

| | REALISATION | | | PREVISION | | |
|---|---|---|---|---|---|---|
| Service A destiné Client | N | N+1 | N+2 | N+3 | N+4 | N+5 |
| Quantité Service A | 3 | 5 | 8 | 12 | 15 | 20 |
| Prix HT Service A | 40,000 | 40,200 | 40,500 | 40,700 | 50,000 | 50,200 |
| *Ventes* Service A | 120,000 | 210,000 | 324,000 | 488,400 | 750,000 | 1,004,000 |
| CHIFFRE D'AFFAIRES GLOBAL | 120,000 | 210,000 | 324,000 | 488,400 | 750,000 | 1,004,000 |

## 3. جدول حسابات النتائج المتوقع:

| العنصر | سيناريو محافظ | سيناريو متوسط | سيناريو جيد |
|---|---|---|---|
| الإيرادات | 700,000دج | 1,500,000دج | 2,800,000دج |
| تكاليف العتاد | 250,000دج | 500,000دج | 900,000دج |
| تكاليف التطوير والاختبار | 400,000دج | 700,000دج | 1,000,000دج |
| التشغيل والصيانة | 100,000دج | 200,000دج | 350,000دج |
| التسويق والتنقل | 50,000دج | 150,000دج | 250,000دج |
| إجمالي التكاليف | 800,000دج | 1,550,000دج | 2,500,000دج |
| النتيجة الصافية | 100,000-دج | 50,000-دج | 300,000دج |

- وضح هذا الجدول أن الربحية في السنة الأولى ليست مضمونة في حال تم الاستثمار في العتاد وتطوير النظام دفعة واحدة، وهو أمر طبيعي في المشاريع الناشئة. لذلك يُفضل اعتماد نموذج تدريجي يبدأ بحجم صغير، ثم إعادة استثمار المداخيل المتحققة من التشغيل في عملية التوسع المستقبلي للمشروع، مما يقلل من المخاطر المالية ويزيد من استدامة المشروع على المدى الطويل.

## 4. خطة الخزينة:

توضح خطة الخزينة تطور التدفقات المالية للمشروع خلال السنة الأولى، من خلال مقارنة المداخيل بالمصاريف الشهرية، بهدف متابعة الوضعية المالية وضمان التوازن بين الإيرادات والنفقات.

في بداية المشروع (الأشهر 1 إلى 2)، تكون المداخيل شبه منعدمة، حيث يركز الفريق على مرحلة التأسيس، والتي تشمل تحليل المتطلبات، تطوير النموذج الأولي، وتجهيز البنية التقنية الأساسية، مما يؤدي إلى تسجيل مصاريف فقط.

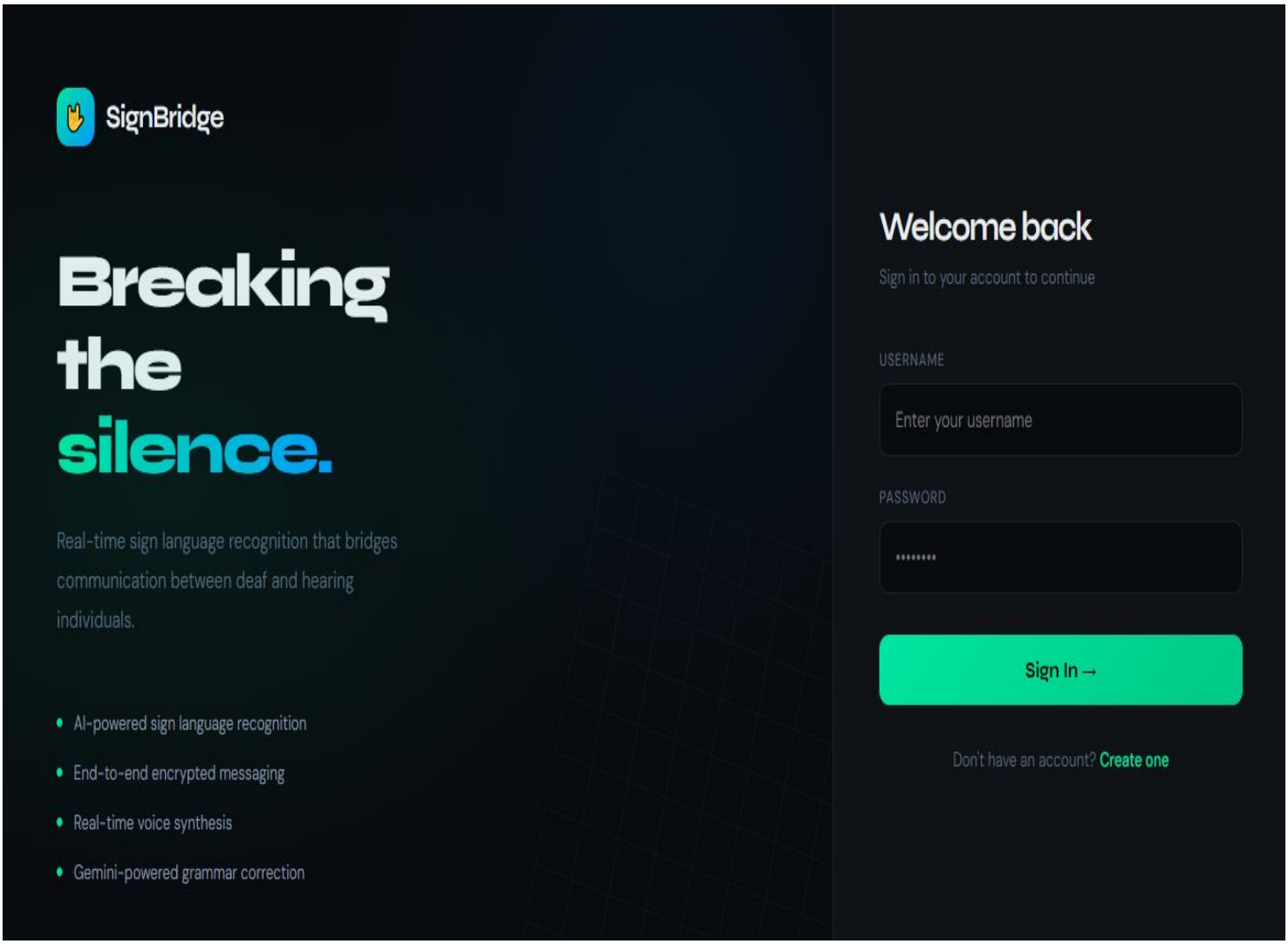
ابتداءً من الشهر الثالث، يبدأ المشروع في تحقيق أولى المداخيل من خلال تجربة النظام مع أول عميل أو شريك، مع استمرار المصاريف المتعلقة بالاختبارات والتحسينات التقنية.

خلال الفترة الممتدة من الشهر الرابع إلى السادس، يبدأ المشروع في تحقيق إيرادات أكثر استقراراً من خلال الاشتراكات وخدمات الصيانة، مقابل مصاريف خاصة بالدعم الفني والتسويق بهدف توسيع قاعدة المستخدمين.

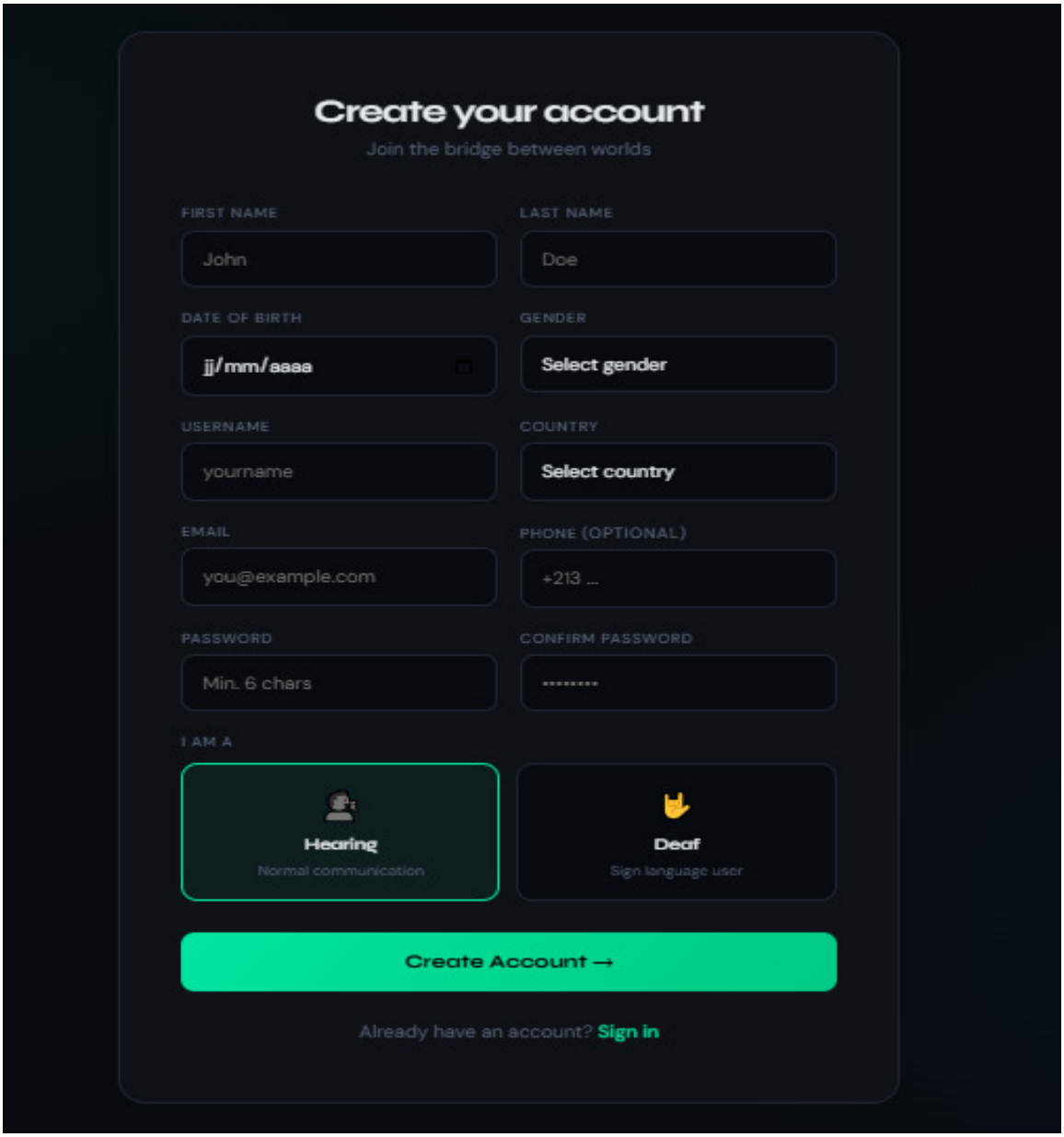
أما في الفترة من الشهر السابع إلى نهاية السنة، فينتقل المشروع إلى مرحلة أكثر استقراراً، حيث تزداد التعاقدات بشكل تدريجي، مع الحفاظ على مصاريف الصيانة والتطوير، مما يسمح بتحقيق توازن مالي ونمو تدريجي في الخزينة.

بشكل عام، تعتمد الخطة المالية للمشروع على مبدأ النمو التدريجي، حيث يتم إعادة استثمار الإيرادات في تطوير وتحسين النظام، مما يقلل المخاطر المالية ويعزز استدامة المشروع على المدى الطويل.

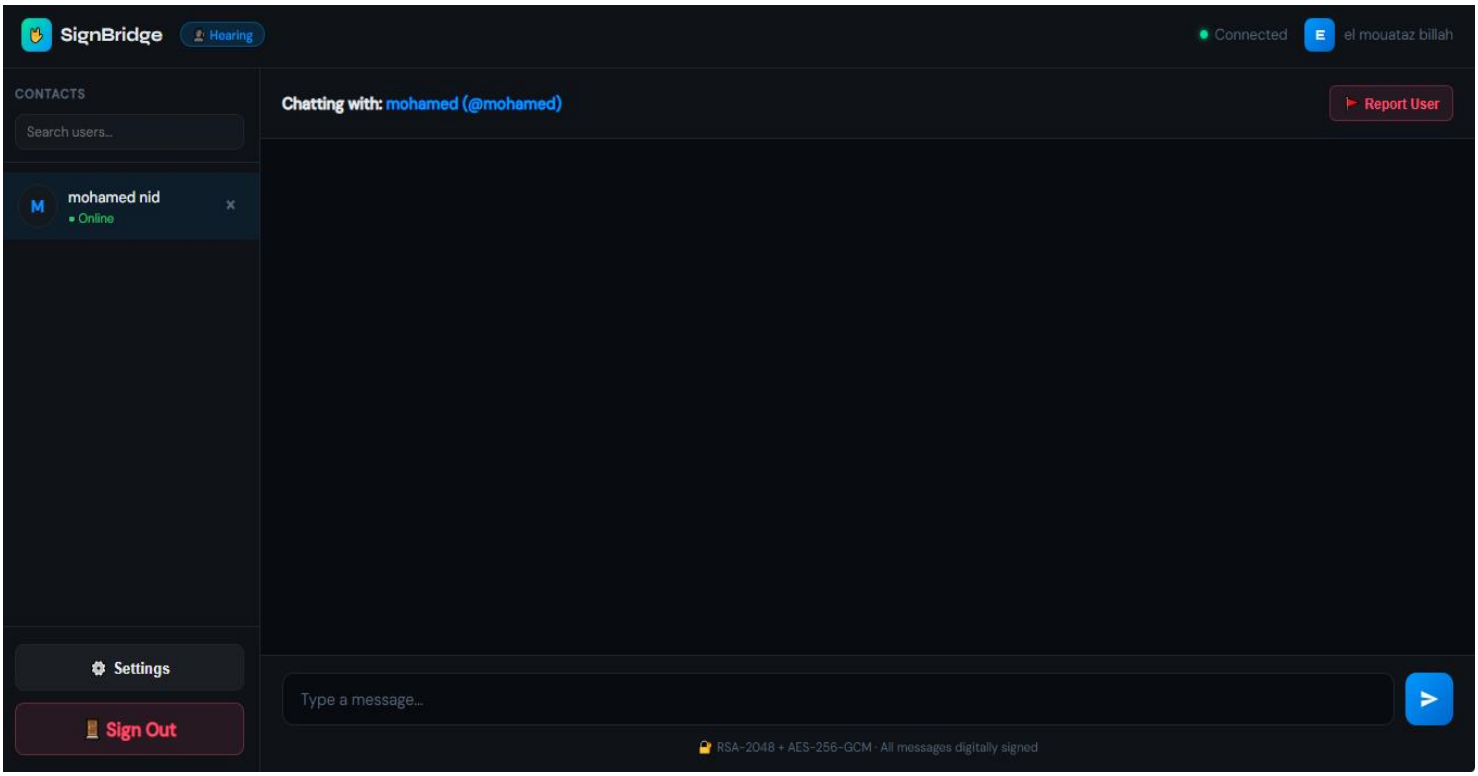
# المحور السادس: النموذج الاولي التجريبي



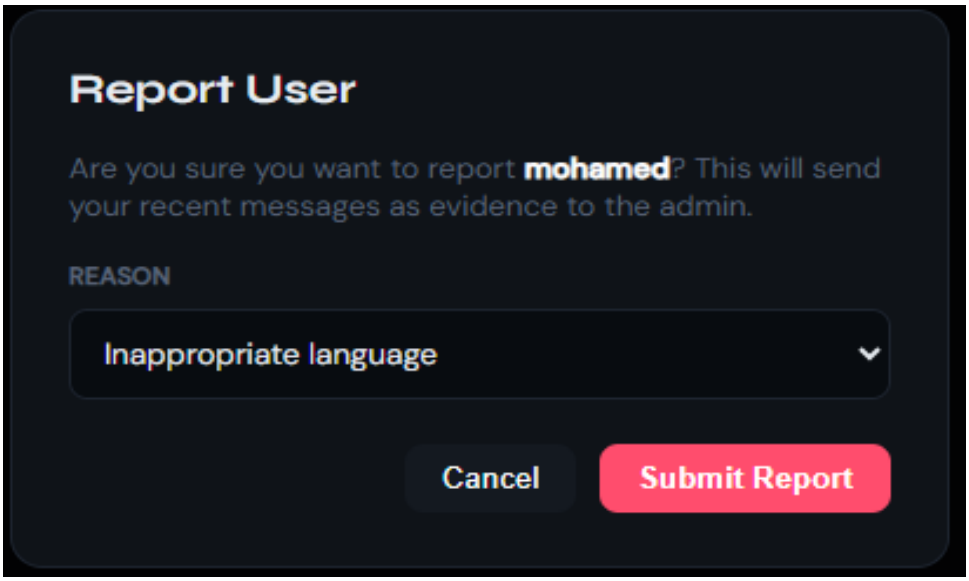

واجهة تسجيل الدخول الى حساب **Sign bridge**




واجهة انشاء حساب في Sign bridge

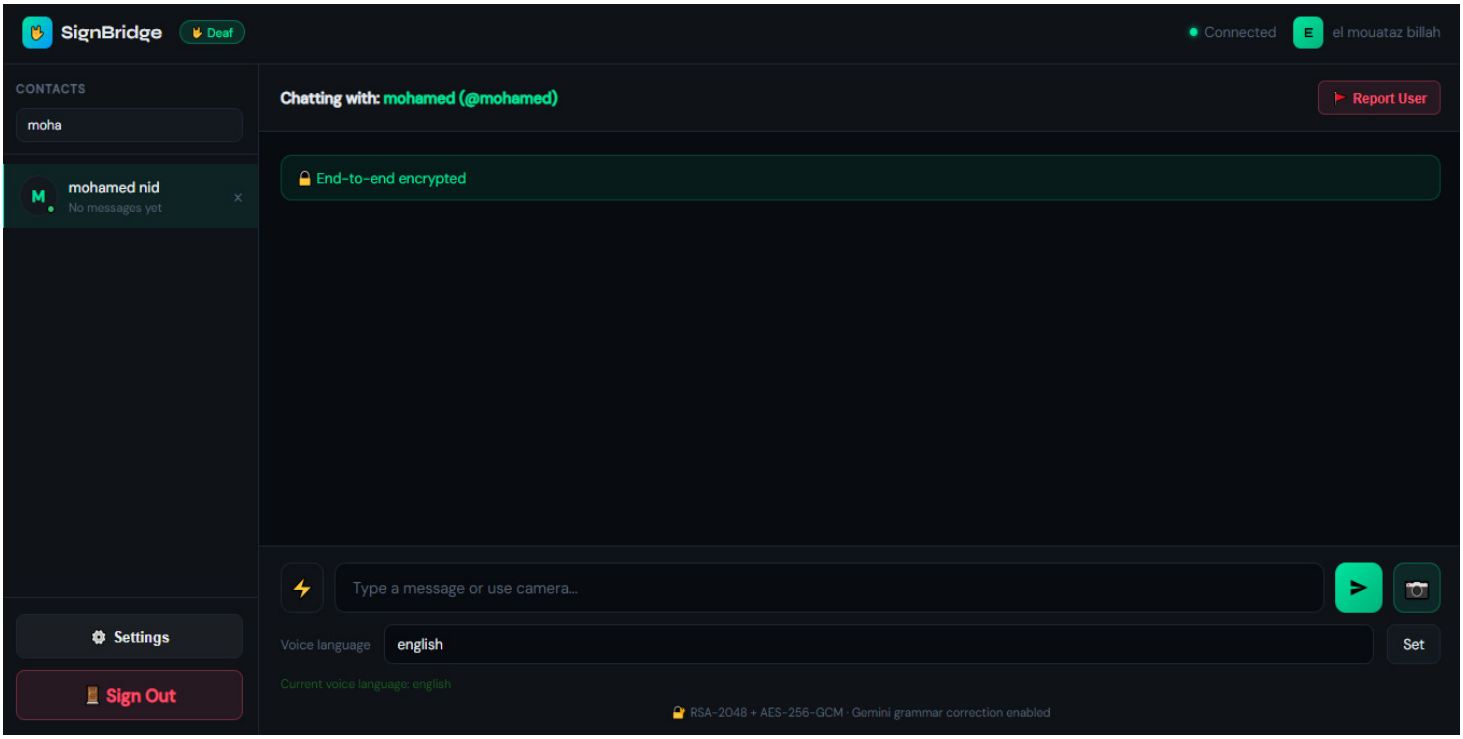



واجهة الدردشة للشخص العادي

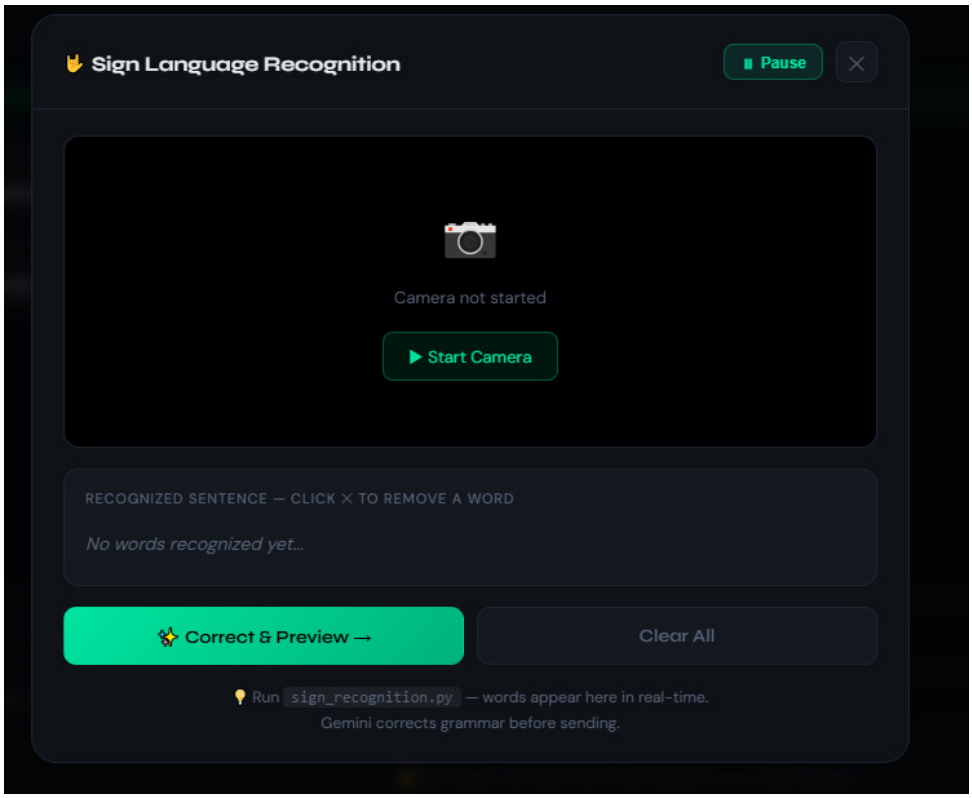



واجهة ارسال التقارير للadmin

مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم
والأشخاص غير القادرين على الكلام

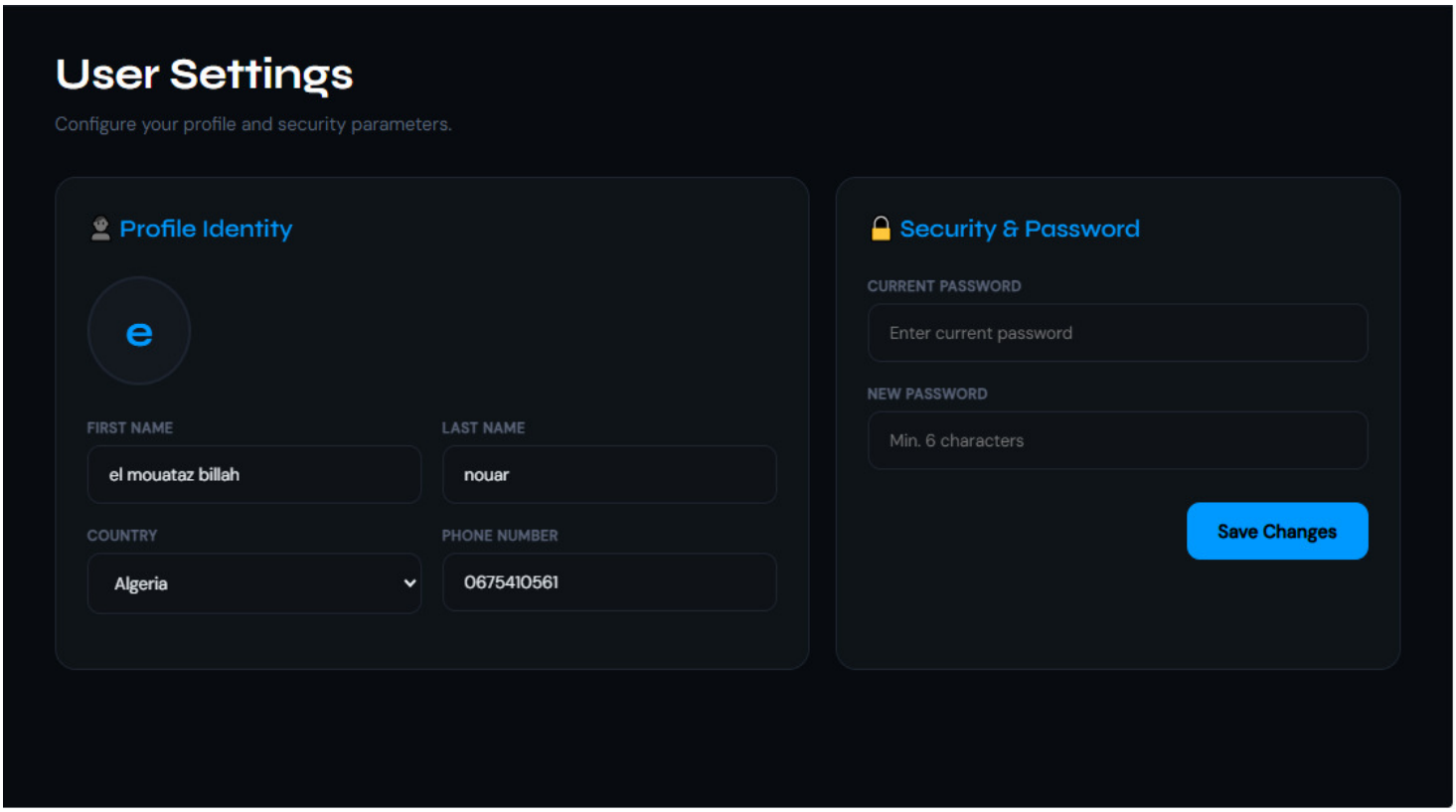



واجهة الدردشة للشخص الاصم /الابكم

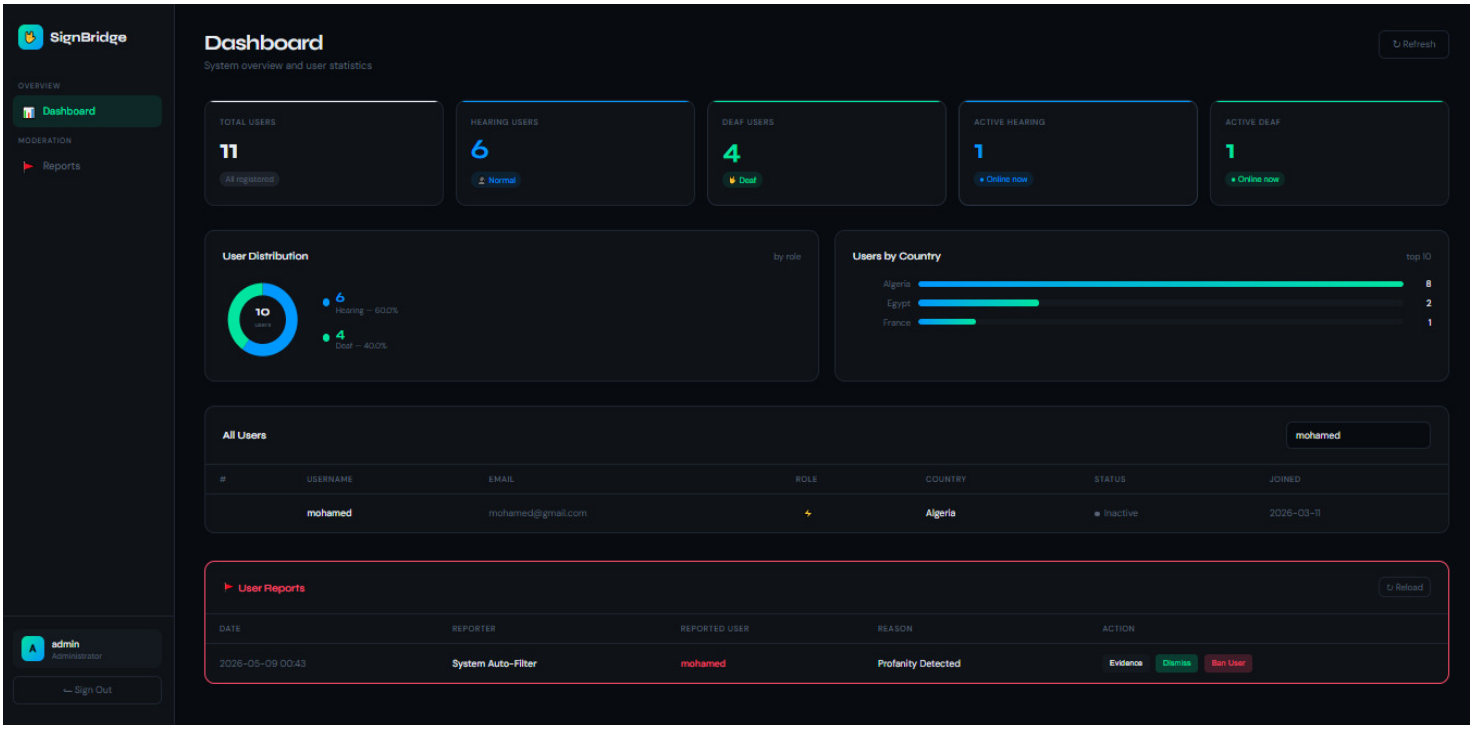



واجهة التعرف على لغة الاشارة للشخص الاصم /الابكم

مساعد تواصل قائم على النماذج اللغوية الكبيرة (LLM) للصم
والأشخاص غير القادرين على الكلام




واجهة الاعدادات للمستخدم




واجهة لوحة التحكم للadmin

# Banking App Business Model Canvas

Sign Language Communication Platform

## Key Partners
الشركات الرئيسية

- جمعيات الصم والمنظمات غير الحكومية الداعمة (الثقة المجتمعية + التبني)
- الحكومات وهيئات تيسير الوصول (التنظيم + التبني على نطاق واسع)
- الجامعات ومختبرات الأبحاث
- مزودي البنية التحتية السحابية والذكاء الاصطناعي (وحدات معالجة الرسومات، الاستضافة، خدمات النماذج)
- منصات الاتصال المرئي (قنوات التكامل للاستخدام في الوقت الفعلي)
- خبراء ومترجمو لغة الإشارة

## Key Activities
الانشطة الرئيسية

- تطوير وتحسين نماذج الذكاء الاصطناعي بشكل مستمر (التدريب، التقييم، والتحسين)
- التعرف على لغة الإشارة في الوقت الفعلي (تحويل الإشارة إلى نص/صوت)
- تحويل النص إلى فيديو بلغة الإشارة (Text-to-Sign) لتسهيل التواصل ثنائي الاتجاه
- تحسين الأداء في الزمن الحقيقي (تقليل زمن الاستجابة، القابلية للتوسع)
- جمع البيانات الخاصة بلغة الإشارة وتوسيمها وإدارتها
- تطوير المنصة الإلكترونية وصيانتها بشكل مستمر
- ضمان حماية البيانات والخصوصية والامتثال لمعايير الأمان
- التحقق من دقة النماذج ومراقبة الأداء بشكل مستمر وتقييمه

## Key Resources
الموارد الرئيسية

- نماذج الذكاء الاصطناعي الخاصة (تحويل لغة الإشارة إلى نص، وتحويل النص إلى لغة الإشارة، ومعالجة الكلام)
- مجموعات بيانات لغة الإشارة
- فريق الهندسة والذكاء الاصطناعي
- البنية التحتية السحابية ووحدات معالجة الرسومات (GPU)
- منصة الويب
- منصات وتقنيات الذكاء الاصطناعي (API وخدمات تدريب النماذج)

## Value Proposition
القيم المقترحة

**Core impact**

- الاستقلال عن المترجمين الفوريين.
- لمساواة في فرص الحصول على التعليم والعمل والرعاية الصحية والخدمات العامة.
- الإدماج الكامل للأشخاص الصم وضعاف السمع وغير القادرين على التعبير اللفظي

**Unique differentiation**

- المنصة الوحيدة التي تتيح التواصل التام في الاتجاهين (بين الصم ↔ السامعين)
- أول نظام للتواصل بلغة الإشارة في الوقت الفعلي وفي الاتجاهين
- مصمم للاستخدام في بيئات الحياة اليومية، وليس للعروض التجريبية المختبرية

**Core functional value**

- اتصال فوري دون أي تأخير ملحوظ
- نص/صوت (وضع الصم)→لغة الإشارة
- فيديو بلغة الإشارة (وضع السمع) → نص

**Accessibility & deployment advantage**

- منصة قائمة على الويب (لا تتطلب تثبيتًا)
- تعمل على مختلف الأجهزة لتوفير إمكانية الوصول الفوري
- مصممة للاستخدام المؤسسي واليومي.

## Customer Relationships
العلاقات مع العملاء

- منصة ذاتية الخدمة مع تجربة إدماج سهلة وبديهية للمستخدمين الجدد
- برنامج سفراء من مجتمع الصم لتعزيز الثقة وتشجيع تبني النظام
- آليات مستمرة لجمع ملاحظات المستخدمين لتحسين نماذج الذكاء الاصطناعي
- دعم مؤسسي مخصص (التدريب وإدارة الحسابات للجهات الشريكة)
- تعزيز التفاعل المجتمعي من خلال الفعاليات والمجموعات التوعوية
- توفير مركز مساعدة وقنوات دعم متعددة (الدردشة، البريد الإلكتروني)

## Channels
القنوات

- منصة الويب (الرئيسية)
- وسائل التواصل الاجتماعي (الموجهة لمجتمع الصم)
- المبيعات المؤسسية المباشرة بين الشركات (B2B)
- 
- العروض التوضيحية وورش العمل للتعريف بالنظام

## Customer Segments
شرائح العملاء

**B2C**

- الأشخاص الصم وضعاف السمع
- لأشخاص غير القادرين على التعبير اللفظي
- الأشخاص السمعيون (الأسرة والأصدقاء وزملاء العمل)

**B2B**

- المدارس والجامعات
- فرق العمل في الشركات وبيئات التعاون عن بُعد
- لشركات التي تتعامل مباشرة مع العملاء (البنوك، الاتصالات، تجارة التجزئة، مراكز الدعم)

## Cost Structure
هيكل التكاليف

- تكاليف الاستدلال بالذكاء الاصطناعي والحوسبة باستخدام وحدة معالجة الرسومات (الاستخدام في الوقت الفعلي، والتوسع)
- تكاليف تدريب نماذج الذكاء الاصطناعي وإجراء التجارب عليها
- تكاليف مسار عمليات جمع البيانات وتصنيفها ووضع العلامات عليها
- رواتب موظفي الهندسة وتطوير المنتجات
- البنية التحتية السحابية (التخزين، والاستضافة، وعرض النطاق الترددي)
- تكاليف الامتثال للوائح التنظيمية وشهادات إمكانية الوصول

## Revenue Streams
مصادر الايرادات

- اشتراك SaaS بين الشركات (ترخيص مؤسسي)
- اشتراك مميز (دقائق غير محدودة + دقة أعلى)
- نموذج "فريميوم" (استخدام/ميزات محدودة)
- التسعير حسب الاستخدام (الترجمة بنظام الدفع بالدقيقة)
- باقات الشركات (حلول مخصصة وعمليات تكامل)