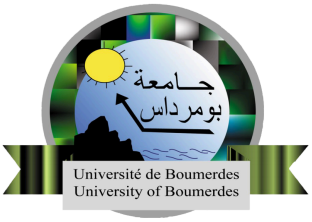
People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

M'Hamed Bougara University - Boumerdes



Faculty of Sciences

Department of Computer Science

Field : Mathematics and Computer Science
Major : Computer Science
Specialty : Information Systems
Specialty Accreditation Order Number : Order No. 872 of 26/07/2016

Final year project submitted in partial fulfillment of the requirements for the

Bachelor's Degree in Computer Science

Title

Design and Development of an Intelligent Chatbot System for Early ASD Screening

Presented by :

ADJERID Wafaa
MERAZGA Abla
MESSADIA Yakoub
NAIT YAHIA Aya

Supervised by: HACHEMI Rania ***Examiner: CHAABANI Mohamed***

Academic Year: 2025/2026

Acknowledgments

We deeply thank Hachemi Rania for her incredible supervision and all the support she has given us as we underwent this project.

We would also like to express our sincere gratitude to the Head of the Computer Science Department for their support and for providing the necessary conditions to carry out this work.

We thank all the teachers that have granted us sufficient knowledge to make this project into a reality.

We express our gratitude to all members of the jury agreeing to evaluate this work.

Finally all thanks to people who have helped and supported us throughout this journey, their support is truly appreciated.

Dedication

I dedicate this project to my family, for their constant love and support.
To my teammates, in appreciation of their effort, collaboration, and dedication throughout this project.
And to myself, for the patience and determination to accomplish this work.

Adjerid Wafaa

This project is dedicated to myself, for my patience, effort, and determination to complete this work despite circumstances and health issues.
To my father, for his guidance, support, and sacrifices.
And to my family and teammates, for their constant encouragement, cooperation, and valuable contribution to this project.

Merazga Abla

First, I thank God for giving me the strength and steadfastness to continue on this path.

To a very special person who holds a dear place in my heart , I am deeply grateful to you for encouraging me to love and improve my level in mathematics, which led me to choose the maths stream in high school and gave me a great opportunity to pursue computer science. Thank you for your presence during those years, for always being there for me, and for supporting me every time. May God keep you happy for your entire life. I hope I have made you proud as I promised you.

To my beloved family my Mother , My father , my brother, and my little sister without forgetting my two aunts thank you for your unconditional support, love, and generosity, and for always being by my side. Thank you for giving my life its meaning and purpose

A special thanks to my older sister, for encouraging me, doing everything to make me happy even during your moments of exhaustion, and for always trying to keep my smile. I love you so much my doctor

To my supervisor Ms. Hachemi Rania, thank you for supervising us and for giving us the support to show our skills in this project. I wish you all the best in your PhD.

A big thank you to all my teachers. I am truly grateful to every teacher from middle school to my bachelor’s degree , everyone who has taught me .
To all the people who have helped me in one way or another, whether in my studies or in other circumstances.

And finally, to myself.
To the me of the past, full of doubts and lacking confidence — we did it !
To the me of the future, I can't wait to see what we will become.

Nait Yahia Aya

Abstract

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by persistent difficulties in social communication and restricted, repetitive behaviors. Early detection is crucial, as timely intervention during childhood can significantly improve developmental outcomes. However, current screening methods are often limited by accessibility, cost, and delays in specialist evaluation.

This work proposes an AI-based screening approach using caregiver-written descriptions of child behavior. The system follows a two-stage pipeline: a conversational large language model that extracts structured behavioral observations from free-text input, followed by a fine-tuned Funnel-Transformer classifier that identifies ASD-related patterns and assigns risk scores. Unlike traditional methods relying on structured questionnaires or multimodal data, this approach focuses on natural language, making it more practical and accessible in real-world settings.

Experimental results show strong performance, achieving 97.08% accuracy, 95.92% precision, 98.6% recall, and a 97.24% F1 score. These results indicate that Funnel-transformer models can effectively capture behavioral signals from text, with particularly high recall helping reduce missed cases. Overall, the proposed system offers a lightweight and scalable tool to support early ASD screening and bridge the gap between parental concern and clinical diagnosis.

Keywords: Autism Spectrum Disorder, Transformer, Text Classification, Healthcare Artificial Intelligence, Clinical Decision Support Systems, Health Informatics

Table of contents

List of Acronyms : **1**
General Introduction **2**
Background and Basic Concepts **3**
1.1 ASD and the importance of early detection 3
1.2 Problem Statement 4
1.3 Objectives 4
1.4 Conclusion 5
State of The Art **6**
2.1 AI and machine learning in ASD detection 6
2.2 Deep learning approaches and relevancy to real world situations 6
2.3 Natural Language Processing and Text-Based Detection 7
2.4 Interactive systems and Chatbots 7
2.5 Research Gap and Motivation 7
Methodology and System Design **8**
3.1 System Overview 8
3.2 Dataset 8
3.3 The Two-Model Pipeline 10
3.3.1 Funnel Transformer Classifier 10
3.3.1.1 Training Configuration and Hyperparameters 10
3.3.1.2 Model Selection: Funnel vs BERT 11
3.3.2 Llama 3.3 agent 12
3.3.3 The pipeline 12
3.4 Experiments And Results 13
Implementation **15**
4.1 Diagrams 15
4.1.1 Use Case Diagram 15
4.1.2 Activity Diagrams 16
4.1.2.1 Login/Sign Up Diagrams: 16
4.1.2.2 Chatbot Interaction Diagram 17
4.1.3 Sequence Diagram 18
4.1.4 Class Diagram 19
4.2 Development Environment 20
4.3 User Interface 21
4.4 Backend server 23
4.5 Database 23
Bibliography **25**

List of Figures

Figure 3.1: Reading Data of TASD dataset……..……8
Figure 3.2: Architecture of the two model based ASD detection framework……13
Figure 3.3: Confusion matrix of funnel classifier……14
Figure 4.1: Use Case Diagram for user interaction with ASDetect Bot……...15
Figure 4.2: Activity Diagram for user Login/Sign-up……..16
Figure 4.3: Activity Diagram for ASDetect Bot interaction……17
Figure 4.4: : Sequence Diagram for user interactions in ASDetect Bot……...18
Figure 4.5: Class Diagram for ASDetect Bot Database ……..19
Figure 4.6: Home page of the ASDetect Bot website……..21
Figure 4.7: Authentication page of the ASDetect Bot website……22
Figure 4.8: Privacy page of the ASDetect Bot website……22
Figure 4.9: Main interface of the ASDetect Bot website…….22

List of Tables

Table 3.1: Dataset Overview……..……9
Table 3.2: ASD labels count ……9
Table 3.3: Hyperparameters used for fine-tuning Funnel ……10
Table 3.4: Comparison between BERT and Funnel Transformer ……11
Table 3.2: Classification performance of the model……..……14
Table 4.1: used softwares and technologies……..……20

List of Acronyms :

ASD — Autism Spectrum Disorder
AI — Artificial Intelligence
NLP — Natural Language Processing
LLM — Large Language Model
ML — Machine Learning
BERT — Bidirectional Encoder Representations from Transformers
GCN — Graph Convolutional Network
API — Application Programming Interface
DBMS — Database Management System
JSON — JavaScript Object Notation
CLS — Classification Token (transformer)
REST — Representational State Transfer

General Introduction

Autism Spectrum Disorder represents one of the hardest challenges in early childhood development for children who have it, not only is it misunderstood, but many children do not get their first diagnosis until much later in their life, regardless of whether parents might notice the signs earlier.

Our project here presents the design and development of ASDetectBot, a screening tool that allows parents to describe their child's behavior and receive a report alongside feedback to support early identification of ASD. The chatbot does not diagnose nor replace the judgement of a trained clinician, but it serves as a tool that detects the signs and refers the parent to a professional if necessary.

The work conducted is organized across 4 chapters as follows:

Chapter 1:Background and basic concepts of Autism Spectrum Disorder (ASD), highlights the importance of early detection, presents the problem statement, and sets objectives.

Chapter 2: Analyzes related literature, relevancy of previous research, research gaps and the projects approach towards them.

Chapter 3: Presents the methodology and system design, including the dataset

Chapter 4: implementation of the system,technologies and tools used, frontend and backend development, diagrams pertaining to the project, and the database design.

Background and Basic Concepts

Autism spectrum disorder (ASD) is a neurodevelopmental condition made up of enduring challenges in social communication and interaction and the presence of repetitive and restricted behavior patterns.

Early diagnosis of Autism Spectrum Disorder is important, yet the current screening methods don't provide the factor of accessibility and accuracy which can delay the identification of this disorder in kids. However, recent advances in AI and Natural Language Processing (NLP) offered tools that have shown decently accurate performance while remaining practical and easy to use especially for parents.

In particular, Transformer-based models have proven themselves to excel in understanding text-based data and extracting information from it. Making them an ideal fit for analyzing the descriptions that the caregivers provide about their child's behavior. This project in particular uses a two-model pipeline: a transformer based-LLM (Llama 3.3 70B) that runs the conversation and extract structured behavioral observation from each response, which will be passed down to a fine tuned Funnel-transformer that is trained on labeled text-based data to classify each observation and assign the risk score to each detected pattern. This system offers an efficient tool that enables our endeavors to support early screening for parents, legal guardians, or anyone in need of the service.

1.1 ASD and the importance of early detection

The high prevalence of Autism Spectrum Disorder (ASD) necessitates acquiring early detection tools that can lead to appropriate interventions for children and assist children in performing as many life functions as possible. These early years of childhood remain those most receptive to intervention, adaptation, and cognitive development:

– "There is an unwritten agreement that this disorder can be minimized by providing early intervention [...] which is preceded by early detection" **[1]**.

Furthermore, early detection tools for ASD can significantly impact the child:
– "Early intervention can improve cognitive and behavioral functioning in children with autism.[...] Fuller EA et al. describe that major improvements were observed in children's social communication when clinicians' early intervention showed that ASD spectrum disorder prognosis could largely be prevented at the earliest stages of children's communication development"**[2]**.

1.2 Problem Statement

A lot of emphasis is put not only on the condition but also time; the interval during which the signs are present, faintly flowering, while recognition arrives too late to seize the full promise of early intervention. Parents notice fragments long before diagnoses are spoken aloud: the absent gaze, the delayed response to a name, the peculiar devotion to repetition. But suspicion alone is not a diagnosis, and certainty remains barricaded behind specialists, lengthy evaluations, financial burden, and tools whose complexity renders them inaccessible to many families.

This presents a growing problem: although early diagnosis has repeatedly been shown to improve developmental, communicative, and social outcomes in children with ASD, access to such diagnosis remains uneven, expensive, and unnecessarily delayed.

1.3 Objectives

The primary objective of this project is to design and evaluate an accessible screening tool that makes conversations and supports an early identification of ASD in toddlers based on their parents description of their behavior. In order to achieve this, we pursue the following objectives:

• Develop a chatbot that is capable of making conversations, understanding context from natural language written text and replying accordingly, which allows the parent to freely describe their child's behavior in an accurate and precise way and make it more engaging to them.

• Classify the parent's reported behavioral observations of the toddler as either ASD if there's consistent signs and patterns, or as non-ASD if the child behavior is typical.

• Ensure that the outputs are clearly framed as screening support rather than a substitute for clinical diagnosis, and encourage the parent to seek professional assessment for their toddler.

• Build an accessible web-based Interface that is accessible, free and doesn't require any technical knowledge to be able to use it

In short: We endeavor for the creation of a simple and accessible screening tool capable of assisting parents in identifying early indicators of ASD based on textual input. Not as a substitute for clinical diagnosis, but a means of shortening the distance between concern and intervention, between hesitation and help.

1.4 Conclusion

A chatbot may assist parents by presenting screening questions, interpreting responses, and identifying potential early indicators of ASD in children. While such systems do not replace clinical diagnosis, they can contribute to earlier awareness and orientation toward specialized evaluation, thereby helping reduce delays in intervention and support.

State of The Art

The early identification of Autism Spectrum Disorder (ASD) is important for timely interventions and support. Recently a number of studies have been conducted in order to apply Artificial Intelligence, Natural Language Processing (NLP), and interactive systems to make early screening work better In this part we will look to earlier research that has looked at AI-based autism detection; NLP model types including transformer based architectures Like Bert and chatbot systems for interactive analysis of parental input.

2.1 AI and machine learning in ASD detection

Many research studies have examined how machine learning can be applied to identifying Autism Spectrum Disorder (ASD) earlier in life through structured data that can include survey responses and observational assessments. Through the implementation of automated machine learning techniques, research has been able to make rapid progress in ASD pattern identification [3]. Other research has attempted to combine the use of parent questionnaire data and video recordings of children playing at home to increase the accuracy of diagnosis [7]. The use of multiple data sources in the aforementioned examples shows that applying multiple modalities to support prediction can improve the performance of the prediction model compared to using one source of data only. However, these examples of research primarily utilize structured input data or multimodal input data and therefore have not been applied to the analysis of unstructured or free text parent descriptions.

2.2 Deep learning approaches and relevancy to real world situations

Deep learning models are increasingly being used to identify complex patterns of behavior related to Autism SpectrumDisorder (ASD). Recently, hybrid architectures, which combine spatial and temporal modeling (for example, using Graph Convolutional Networks (GCNs) combined with sequence-based models), have been shown to achieve very high accuracy at
detecting ASD through analyzing behavioral interactions [7]. Furthermore, improved transparency and accuracy of model predictions can be achieved through the use of explainable artificial intelligence (AI) techniques, and by conducting interaction analysis through analyzing home video footage to increase interpretability and improve the overall prediction accuracy [6]. However, many of these approaches are not practical for use in
most real-world environments, as they tend to require large amounts of complex data (such as video footage) to be analyzed.

2.3 Natural Language Processing and Text-Based Detection

The area of natural language processing has grown as an attractive research opportunity for the analysis of language, and there are a range of studies that have looked at the application of speech transcripts and written documents to detect ASD while also considering privacy [10]. Many researchers have used NLP to extract relevant biomedical concepts from the literature related to autism, which shows how NLP can add value through structuring unstructured text for downstream uses [9]. More recently, the combination of machine learning, deep learning, and NLP (i.e., transformer-based models like BERT) has been successful in the analysis of very large amounts of textual data (e.g., social media), specifically through identifying linguistic patterns associated with ASD [6].

2.4 Interactive systems and Chatbots

Interactive systems have been suggested to support early screening for Autism Spectrum Disorder (ASD) through the use of AI-based tools that can analyze a variety of behavioral signals, such as facial expressions, eye gaze, and physical actions, using both machine learning (ML) and deep learning algorithms [3]..More recently, the emergence of large language models (LLMs) has driven the growth of chatbots by allowing contextualized interactions and greater naturalness of conversation — although barriers still exist with respect to both privacy and trustworthiness. These new technologies create more accessible channels of care with greater levels of user engagement, many rely on complex, multimodal datasets that are difficult for scientists, practitioners, and families to manage.

2.5 Research Gap and Motivation

Despite advances, ASD detection with AI still has many limitations. Most previous systems require structured data (e.g., using a user’s questionnaire) or complex inputs (e.g., videos).

In addition, the use of conversational interfaces combined with advanced NLP models to analyze parents' free-form input in everyday environments has not been effectively explored.

The goal of our body of work is to fill in the above gaps in previous research; analyzing the text descriptions provided by parents in order to support early ASD identification in a practical,easy, and intuitive way; this being our primary motive for using funnel-transformer model instead of regular transformer-based ones, lightweight , faster, and just as efficient and powerful — for the sake of providing accessibility and reducing computational power.

Methodology and System Design

3.1 System Overview

ASDetectBot is a chatbot built as a two-model pipeline. A transformer-based large language model LLM LLama 3.3 70B that is accessed via Groq API, it manages the conversations held between the bot and the user and extract the structured behavioral observations from each response it gets, these observations are sent to our fine tuned pre-trained Funnel-transformer (which had been pre-trained on a dataset of 2054 textual description) to get classified, this classifier gives each observation a probability score that indicates the likelihood of a behavioral pattern that is associated with ASD. And at the end it aggregates these scores into our final screening report indicating the total number of observations and the overall risk level -LOW, MODERATE, HIGH- alongside an appropriate referral recommendation.

3.2 Dataset

The dataset used in this project is the **TASD-Dataset: Text-based Early Autism Spectrum Disorder Detection Dataset for Toddlers [13]**, which focuses on early screening of Autism Spectrum Disorder (ASD) using textual behavioral descriptions. The dataset provides textual descriptions of the toddler's behavior, each annotated with a binary label of whether it can be categorized as ASD (1) or non-ASD(0) , alongside the behavioral class (Figure 3.1, Table 3.1 and Table 3.2).
We augmented the dataset using ChatGPT to increase data variability and compensate for the limited amount of real-life data.

Figure 3.1: Reading Data of TASD dataset

	Text	Class	Sign	ASD
0	My daughter smiled and reached for her toy whe...	Normal behavior	+	Non_ASD
1	My son avoids looking at me when I try to talk...	Poor eye contact	-	ASD
2	Both children laughed while playing hide and s...	Normal behavior	+	Non_ASD
3	My child keeps flapping his hands repeatedly w...	Repetitive behaviors	-	ASD
4	My son pointed at the cookie jar and said cook...	Language	+	Non_ASD

The overall structure and distribution of the dataset is presented in Table 3.1 as follows:

Table3.1: Dataset Overview

Feature	Description
Text	Behavioral description
Class	Behavioral category
Sign	Observed behavioral indicator
Label	ASD (1) / Non-ASD (0)
Metric	**Value**
Total Samples	2054
ASD Samples	1097
Non-ASD Samples	957
Training Set	1643 (80.0%)
Validation Set	411 (20.0%)

ASD classes varied according to the different observed behavioral categories ensuring greater diversity and coverage of ASD manifestation, as shown in Table 3.2

Table 3.2: ASD labels count

ASD Label	Count
Repetitive behaviors	191
Social withdrawal	152
Poor Eye Contact	148
Restricted interests	145
Language Delay	144
Emotional Reaction	95
Change Reaction	47
Possible ASD	47
Cognitive development	46
Emotional Empathy	36
Focused Attention	30
Attention Response	29
Follow Pointing	28
Word Repetition	28
Noise Sensitivity	25
Finger Movements	23
Eye Contact	18
Repetitive Behavior	17
Sharing Interest	15
Toy Arranging	15
Sign Communication	12
Tiptoe Flapping	8
Language	4
Language development	2
Sensory Sensitivity	1

3.3 The Two-Model Pipeline

3.3.1 Funnel Transformer Classifier

Our pipeline starts with a fine-tuned Funnel Transformer model which will classify individual behavioral observations as ASD and non-ASD. We chose this model specifically among other transformer models because it's lightweight thanks to the way its encoder layers function as it compresses the sequence representation progressively through each layer using a pooling mechanism (the same way a funnel does) which reduces computational cost while preserving the information. The model processes the tokenized numerical input through the encoder and it produces an embedding from the CLS token which becomes our contextual representation. This embedding gets passed to a linear classification layer that maps it into two outputs (ASD and non-ASD). Then a softmax layer converts scores into probabilities which gives us confidence scores and an ASD probability for each observation.

3.3.1.1 Training Configuration and Hyperparameters

Table 3.3: Hyperparameters used for fine-tuning Funnel

Hyperparameter	Value	Explanation
Max sequence length	64 tokens	It sets the maximum number of tokens the model processes per input
Number of Epochs	3	It represents the number of times the model will get to pass through the whole training dataset
Batch size	8	The number of samples that gets processed together in one step
Learning rate	0.00002	It represents how fast the model updates its weights during the training
Optimizer	Adam	The algorithm we used to update the weights
Loss function	CrossEntropyLoss	This measures the error between predicted labels and real labels
Classification head	Linear(758, 2)	It adds a linear layer that converts learned features into 2 outputs (ASD/Non-ASD)

The base architecture that we chose to fine-tune on our ASD dataset for classification is funnel-transformer/small which is a pre-trained Transformer that we loaded from Hugging Face, therefore it does not need to learn everything from scratch and we only need slight adjustment to the weights for our specific task. That's exactly why we chose 3 as the number of epochs, using any higher number is going to risk overfitting especially with the small amount of data we have (table 3.1)

3.3.1.2 Model Selection: Funnel vs BERT

BERT and Funnel are both transformer-based models that understand text, but they have differences, mainly in efficiency and how they process sequences:

-BERT keeps the full sentence length at every layer which makes it more accurate with less loss of information but heavier.

-Funnel Transformer compresses the sentence step by step keeping the most important information, which makes it faster and lighter and costs less.

Table 3.4: Comparison between BERT and Funnel Transformer

Feature	BERT	Funnel Transformer
Architecture Type	Standard transformer encoder	Hierarchical compressed transformer
Sequence Handling	Maintaining full sequence length for all layers	Gradually reducing sequence length through pooling
Computational Cost	High	Low
Memory Usage	High memory consumption	Low memory consumption
Speed	Slower	Faster
Representation Detail	Preserves details throughout the network	Slight loss of detail due compression
Performance	Very strong especially for NLP tasks	Strong, efficient but slightly less accurate

3.3.2 Llama 3.3 agent

Llama 3.3 70B accessed via Groq API is a transformer-based LLM that we are using in order to maintain conversations and discussions with the users to get information from them rather than depending on questionnaires (which is what other tools do). The parent starts the conversation by describing their toddler's behaviors that they're concerned about, the model then tracks the domains that had been addressed (such as communication, repetitive behaviors, sensory sensitivities.. etc) and ask the user follow-up questions for the areas that hadn't been mentioned by the caregiver yet. Each time, Llama 3.3 extracts the observations from the user's text and makes them into structured JSON format. These observations get deduplicated then passed to our Funnel classifier which returns the label and the confidence score.
Once the exchange ends, the system generates a screening report.

3.3.3 The pipeline

The overall pipeline, illustrated in Fig. 4.1, consists of 4 stages:

- **Conversation**
- **Text preprocessing and tokenization**
- **Feature extraction**
- **Classification and report**

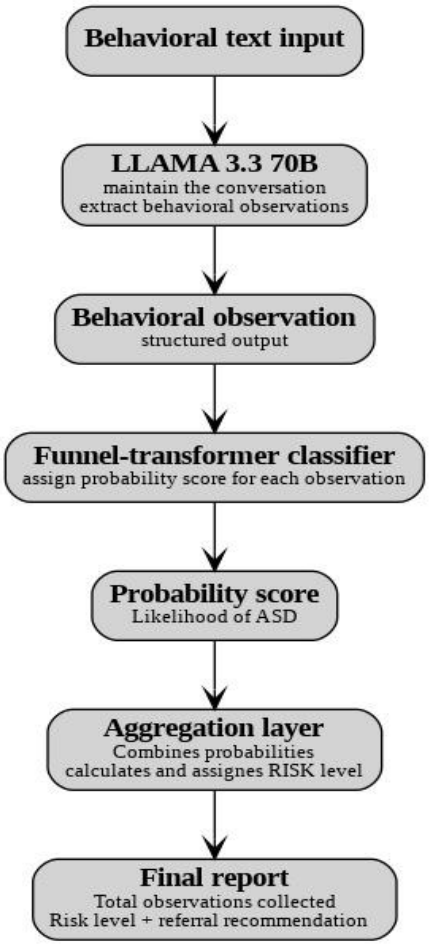


Figure 3.2: Architecture of the two model based ASD detection framework

This architecture allows the model to understand and generate natural language (Llama LLM) while handling the classification accurately (fine-tuned Funnel).

3.4 Experiments And Results

The Funnel transformer classifier was tested on a validation set using standard classification metrics. It performed strongly across all measures, including accuracy, precision, recall, and F1-score. The model achieved an accuracy of 97.08%, with a precision of 95.92% and a recall of 98.60%, resulting in an F1- score of 97.24%. These results indicate that the model is highly effective in identifying ASD-related behavioral patterns from text-based data.

Table 3.5: Classification performance of the model

Metric	Score (%)
Accuracy	97.08
Precision	95.92
Recall	**98.60**
F1-score	97.24

The model achieved a high recall of 98.60% which signifies low rate of false negatives, meaning it catches nearly all potential ASD cases. In screening, this matters a lot because missing a child who actually has ASD and needs support has real consequences. The confusion matrix down below in figure 4.2 confirms that the model correctly identifies most ASD and non-ASD samples, with only a small number of misclassifications.

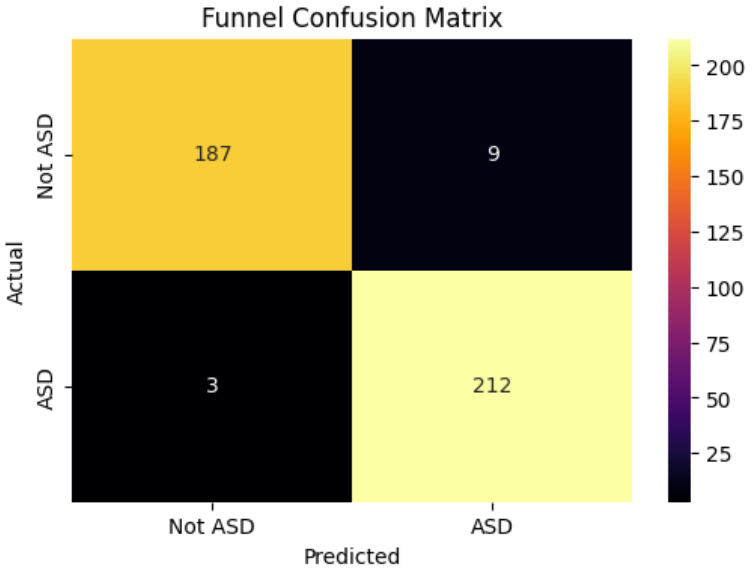


Figure 3.3: Confusion matrix of funnel classifier

Implementation

4.1 Diagrams

4.1.1 Use Case Diagram

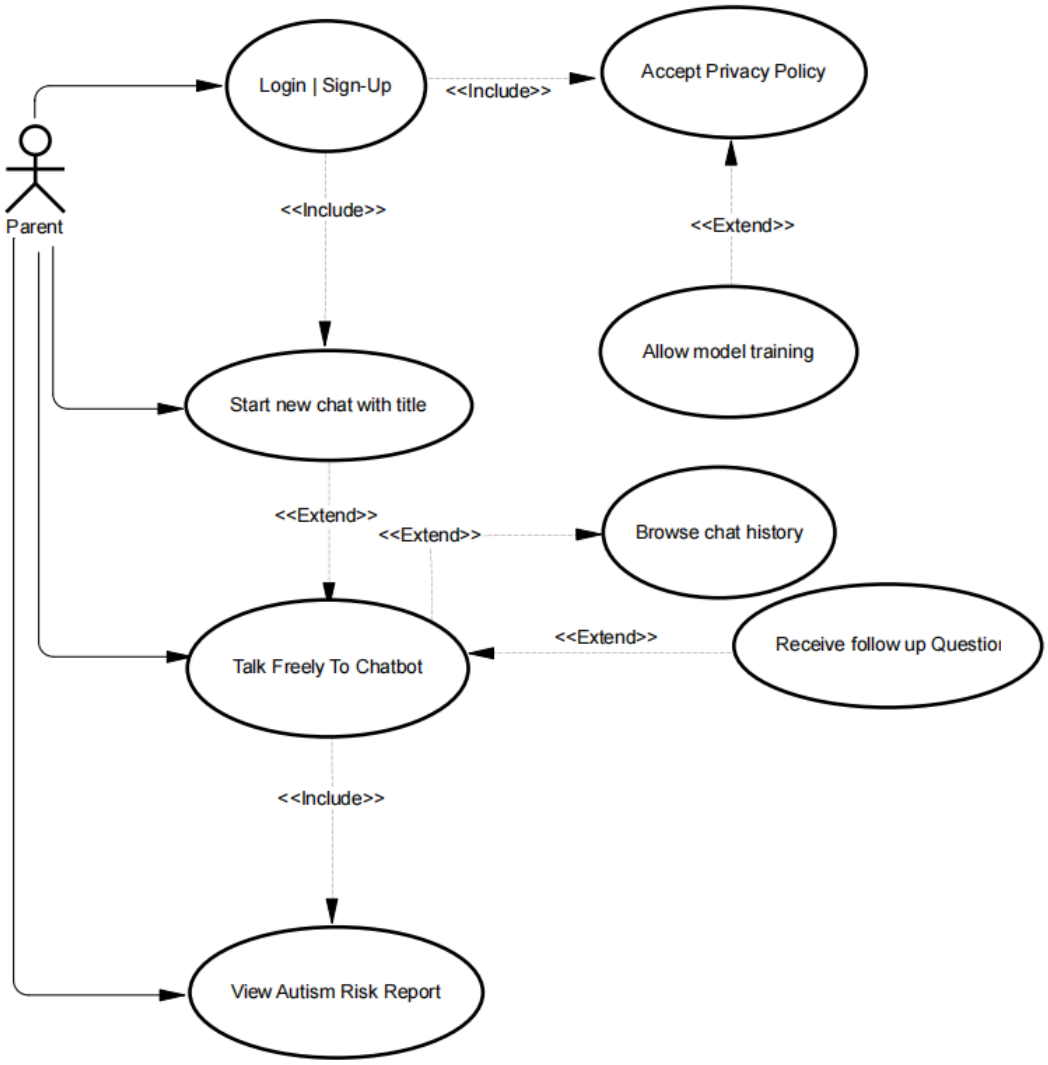


Figure 4.1: Use Case Diagram for user interaction with ASDetect Bot

4.1.2 Activity Diagrams

4.1.2.1 Login/Sign Up Diagrams:

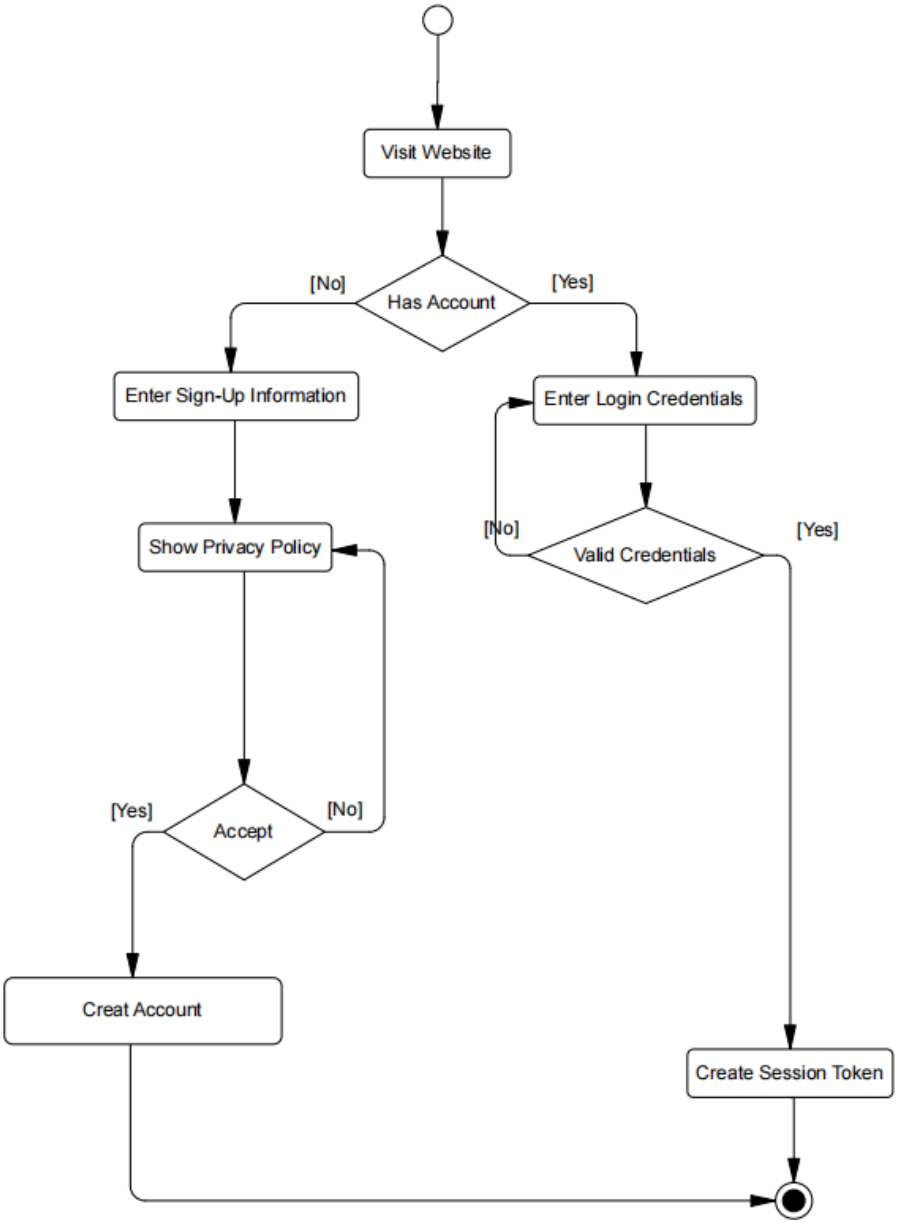


Figure 4.2: Activity Diagram for user Login/Sign-up

4.1.2.2 Chatbot Interaction Diagram

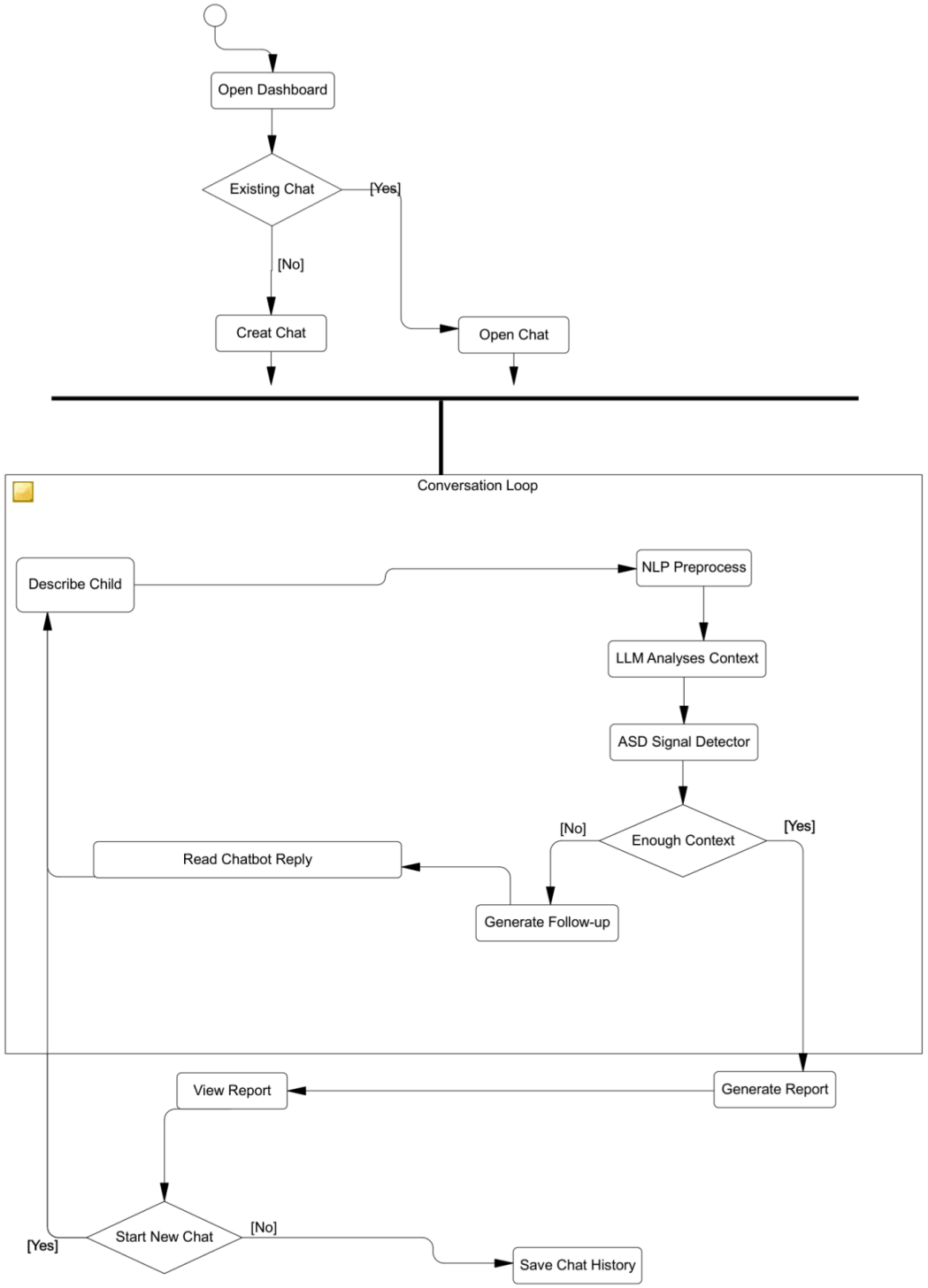


Figure 4.3: Activity Diagram for ASDetect Bot interaction

4.1.3 Sequence Diagram

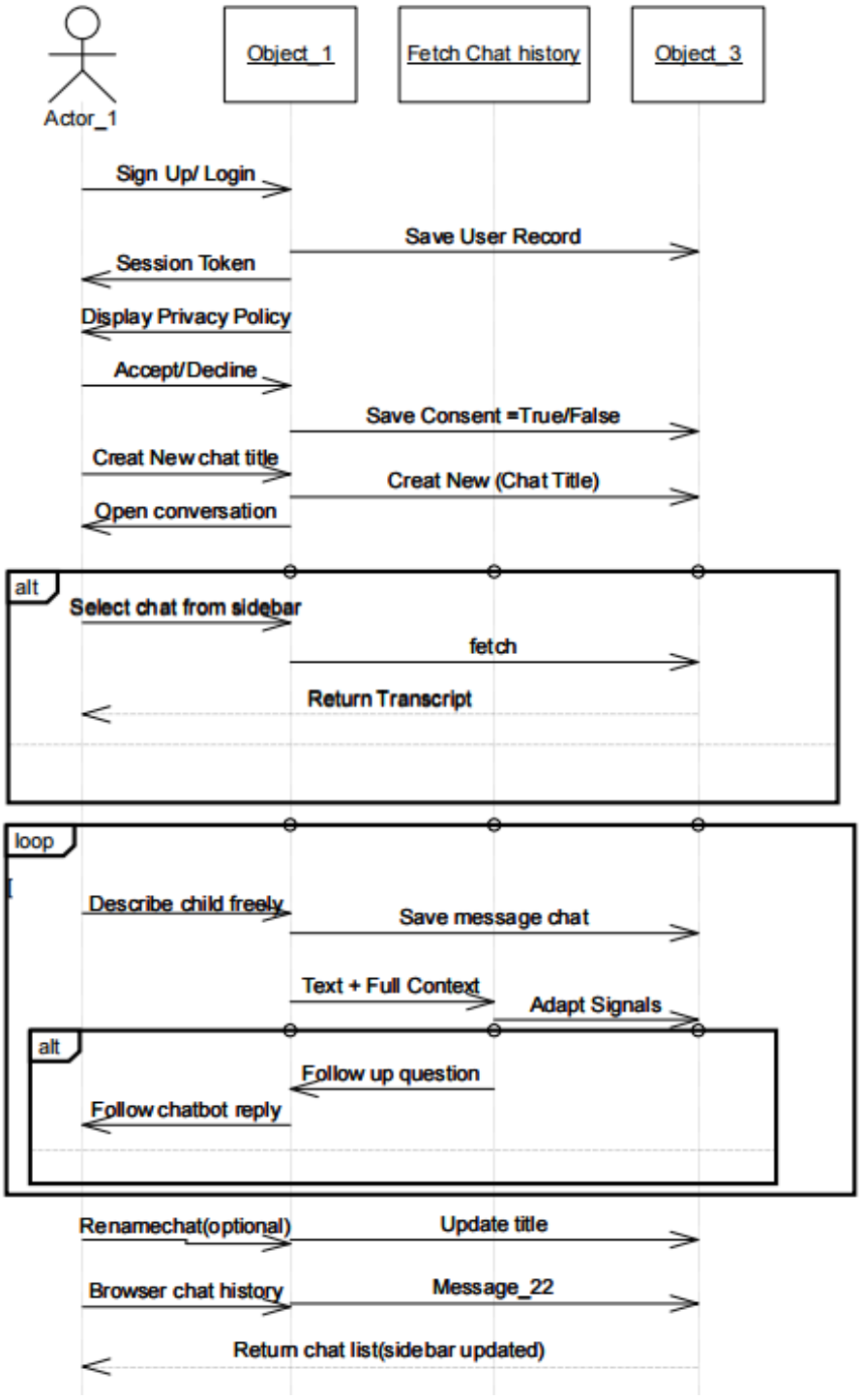


Figure 4.4: Sequence Diagram for user interactions in ASDetect Bot

4.1.4 Class Diagram

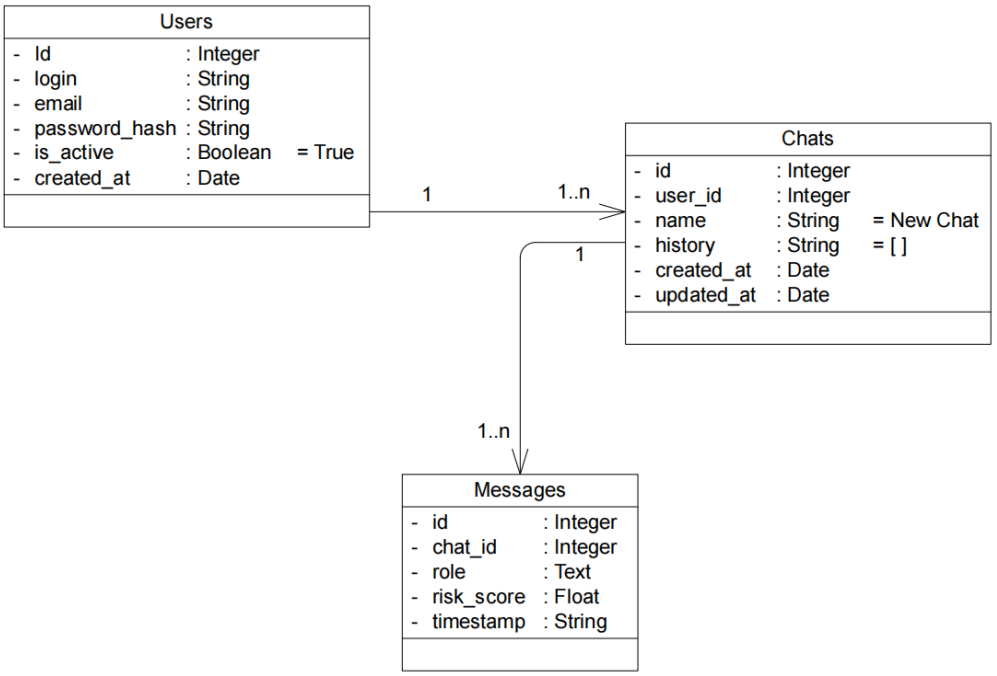
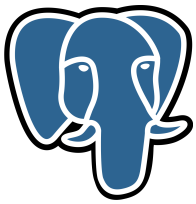


Figure 4.5: Class Diagram for ASDetect Bot Database

4.2 Development Environment

Below is a list of the main software and technologies utilized during the development of the project, managing both the frontend and backend components:

Table 4.1: used softwares and technologies

Software	Definition
	**Visual Studio Code** is a code editor used for writing and managing the source code of our project
	**Google Colab** is a free, cloud-based environment. It allows users to write and execute Python code through the browser, especially suited for data analysis and machine learning because it offers access to GPUs.
	**Flask** is a micro web framework written in Python. Flask supports extensions that can add application features as if they were implemented in Flask itself; it was used for the backend.
	**PostgreSQL** is Open-source relational DBMS that is used for storing and managing the users' data.

4.3 User Interface

The user interface was built using HTML, CSS, and JavaScript. It consists of four pages that the user can navigate as shown in Figure 4.1 and 4.4, and which are:

Home page: It contains a short description of the tool and an input field where the parent can type their first message before being redirected to the chat page if they were already logged in, else they would be redirected to the Sign up page.

Authentication page: It contains two tabs: Login and Sign Up where the user can make a new account after agreeing to the privacy policy or log into their existing account.

Main interface: It contains a messaging area where the conversation takes place, on its left there's the history bar where the user can find all the previous chats. Here the user can create new chats, delete old ones and send messages.

Privacy policy page: It contains an explanation of how the user's data will be stored and used.

All these pages support a dark mode that the user can toggle at any time.

This interface is intuitive and easy to use especially for people that are not familiar with technology, as shown in the figures below:

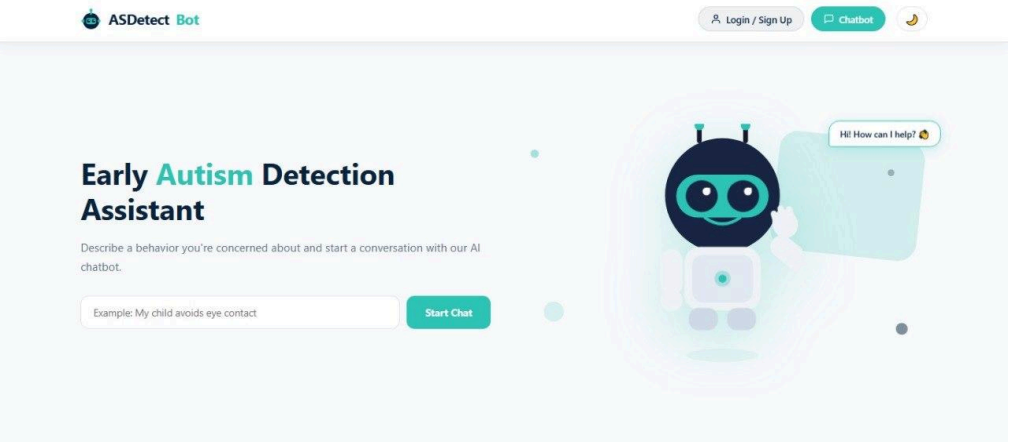


Figure 4.6: Home page of the ASDetect Bot website

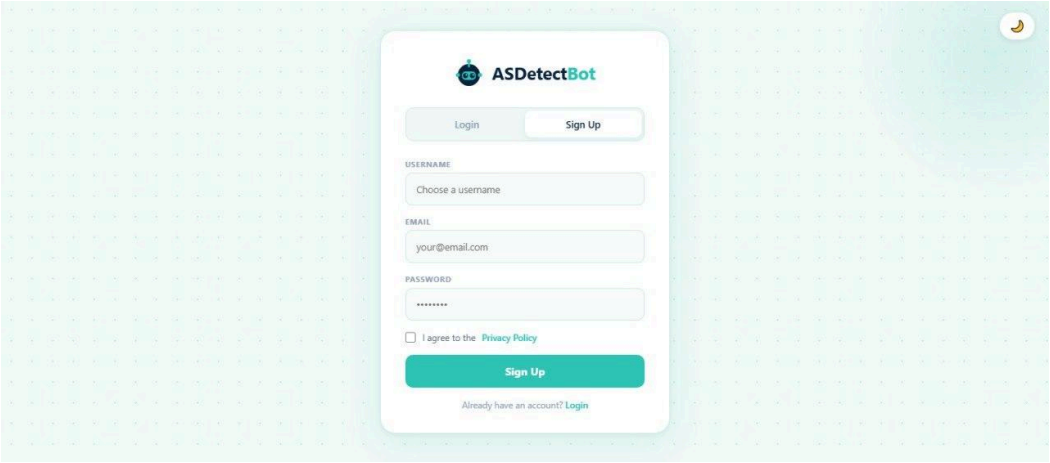


Figure 4.7: Authentication page of the ASDetect Bot website

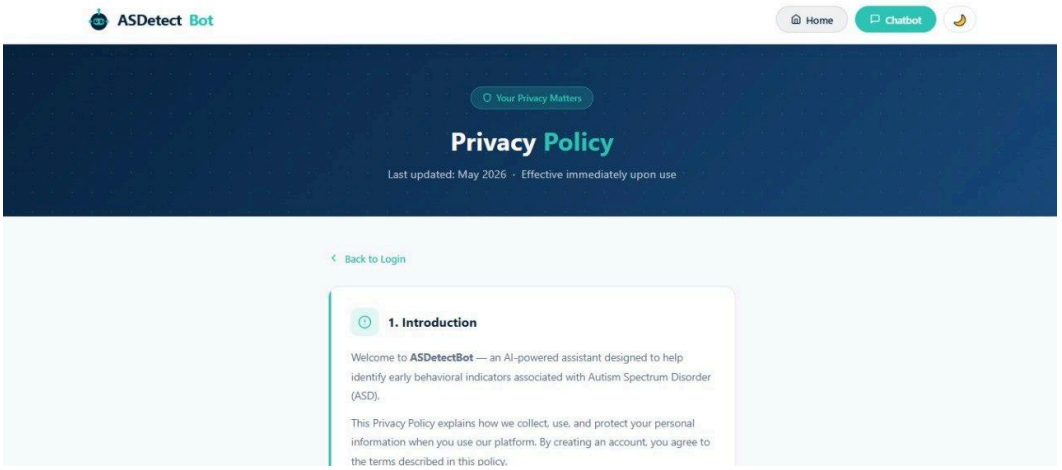


Figure 4.8: Privacy page of the ASDetect Bot website

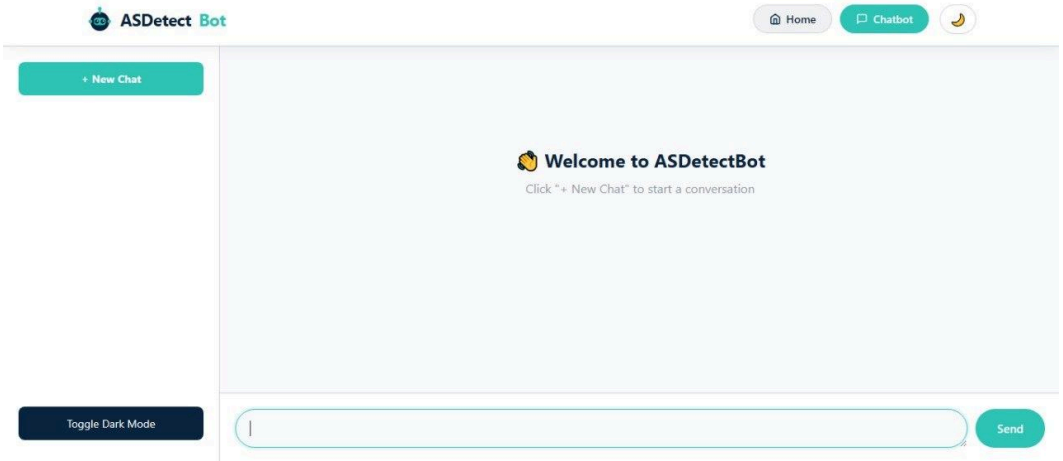


Figure 4.9: Main interface of the ASDetect Bot website

4.4 Backend server

The backend was built using Python Flask which is a framework that set up routes and handle web requests and it serves both the HTML pages and the REST API that the frontend calls.
The authentication is handled using Flask sessions which stores the user ID after login, and all the routes check if the session is valid before processing the user's request, as shown in Figure 4.2 and 4.3.
Users passwords are protected and they're never stored in the database as plain text, instead they're hashed with Bcrypt to ensure the user's security.

4.5 Database

In this project PostgreSQL was used as the database management system. It runs as a separate server and Flask connects to PostgreSQL .

Flask-SQLAlchemy takes care of every task, from querying, inserting, and even updating rows.

The database contains three tables:

users: it stores the parent accounts. Each user has an ID, a username, an email address, a hashed password, and a creation date.

chats: it stores one row per screening session, each chat belongs to a user, has a name, and contains a history column that stores the full Llama conversation as a JSON string.
messages: stores every message that was sent during a session, each message has a role (user or bot), the text content, and a risk score column.

The relationship between these tables is demonstrated in Figure 4.5.

Conclusion

In this project, we designed and implemented ASDetectBot which is a web-based screening tool made to support early ASD screening and to assess the risks in early ages. This tool was made to help parents who notice atypical behaviors and patterns in their kids and are hesitant if it's worth taking a formal assessment or not.

The system follows a structured pipeline that covers data preprocessing, model selection and fine-tuning, backend development, database design, and frontend implementation. It was designed to be reliable as well as usable in real-world settings.

The system combines two AI models, a transformer based LLM Llama 3.3 70B which handles the conversation with the caregiver and gather the information about the toddler by asking questions that covers behavioral domains associated with ASD, and a fine tuned Funnel transformer classifier which was trained on a dataset of 2054 labeled behavioral description which analyzes the observations extracted from the conversation with the parent and assign a risk score. These two models work together to transform the natural language textual input given by the parent into a screening report with a risk level and a referral recommendation. All this runs on a Flask backend and postgreSQL database, and presented with a JavaScript frontend into an interface that is simple and easy to use and navigate around.

ASDetectBot stands out in the way it allows the caregiver to describe the child in a natural conversation which gives more accurate information about the situation and suit the complex nature of human behavior rather than seeing it as a black/white matter in the same way traditional questionnaires with binary input (yes/no) do, while maintaining the simplicity, ease of use and efficiency that makes it accessible for most families and parents.

In conclusion, this work demonstrates how modern natural language processing and transformer-based classification can be combined into a practical and usable tool that helps ASD screening, eases the parents' concern, and encourages them toward taking a formal assessment in the future, which means more children with ASD could get support earlier and avoid the consequences of late diagnosis.

Bibliography

[1] Discourse and Communication for Sustainable Education,vol. 13, no. 2, pp. 13-25, 2022 *Early Detection Assessment Tools in Children With Autism Spectrum Disorder: A Literature study*
[2] Fuller EA, Kaiser AP: The effects of early intervention on social communication outcomes for children with autism spectrum disorder: a meta-analysis. J Autism Dev Disord. 2020, 50:1683-700. 10.1007/s10803-019-03927-z
[3] Design and Verification of an AI-Based Interactive Tool for Early Screening Autism in Children
[4] Detection of Autism Spectrum Disorder (ASD) from Natural Language Text using BERT and ChatGPT Models
[5] Early Detection of Autism Spectrum Disorder Through Automated Machine Learning
[6] Early Prediction of Autism Spectrum Disorders through Interaction Analysis in Home Videos and Explainable Artificial Intelligence
[7] Enhancing Autism Spectrum Disorder Early Detection with the Parent-Child Dyads Block-Play Protocol and an Attention-Enhanced GCN-xLSTM Hybrid Deep Learning Framework
[8] Enhancing ASD Detection Accuracy: A Combined Approach of Machine Learning and Deep Learning Models with Natural Language Processing
[9] Machine Learning Approach for Early Detection of Autism by Combining Questionnaire and Home Video Screening
[10] Natural Language Processing (NLP) Tools in Extracting Biomedical Concepts from Research Articles: A Case Study on Autism Spectrum Disorder
[11] Screening Autism Spectrum Disorder in Children Using Machine Learning on Speech Transcripts
[12] Simulation and Modeling of Expert System for Prediction of Autism using Naïve NLP Method
[13] HACHEMI, Rania; degha, Houssem Eddine (2023), “TASD-Dataset: Text-based Early Autism Spectrum Disorder Detection Dataset for Toddlers”, Mendeley Data, V1, doi: 10.17632/87s2br3ptb.1