School of Computing
FACULTY OF ENGINEERING AND
PHYSICAL SCIENCES

UNIVERSITY OF LEEDS

Final Report

Development and Implementation of a Digital Health AI Prototype for Identifying Autism Spectrum Disorder in Adults

Ayesha Rahman

Submitted in accordance with the requirements for the degree of BSc Computer Science with Artificial Intelligence

2023/24

COMP3931 Individual Project

The candidate confirms that the following have been submitted*:*

Items	Format	Recipient(s) and Date
Final Report	*PDF file*	*Uploaded to Minerva (07/05/24)*
Scanned participant consent forms	*PDF file*	*Uploaded to Minerva (07/05/24)*
Link to online code repository	*URL to GitHub Repository*	*Sent to supervisor and assessor (07/05/24)*
User Guide manual	*URL to GitHub wiki*	*Sent to supervisor and assessor (07/05/24)*
Video Demonstration Link	*URL to YouTube*	*Sent to supervisor and assessor (07/05/24)*

The candidate confirms that the work submitted is their own and the appropriate credit has been given where reference has been made to the work of others.

I understand that failure to attribute material which is obtained from another source may be considered as plagiarism.

(Signature of student) Ayesha Rahman

Summary

This project serves as a vital illustration of how Artificial Intelligence (AI) is becoming an asset in the healthcare field, particularly in the detection of Autism Spectrum Disorder (ASD) diagnosis and management for adults. This has traditionally been used in ASD research and diagnostic criteria, which usually centred on children [3]. The project uses machine learning and model training to help identify autistic characteristics. This tool is aimed at bridging the gap in ASD diagnostics by providing detection and continuous monitoring capabilities.

The project is done in two phases. The first phase was a group effort, involving working on an application framework (LARKS) originally made by the student cohort of the previous year. Under the supervision of Professor Owen Johnson, we decided to refine this framework, improving its features and expanding its capabilities to serve as a sturdy platform for each individual projects.

Upon entering Phase 2, I focused on research and development, with the use of a machine learning model. The objectives included exploring effective methods for ASD detection, developing a machine-learning-based camera tool to analyse eye-movement patterns from video inputs. Another aspect of the project is the integration of a feature that analyses written expressions for signs of ASD in writing, recognizing that such characteristics in adults may present differently than in younger individuals and thus require a tailored approach for detection. Additionally, I implemented an interactive social engagement tool using Natural Language Processing (NLP) to simulate real-life interactions, which alongside the questionnaires (AQ, AQ10, CAT-Q, RAADS-R), helps improve the application. This ensures comprehensive user feedback, allowing users to understand their behaviours and traits in relation to ASD.

Building on this foundation, the project also emphasizes the importance of adaptability and continuous improvement in healthcare technologies. As the understanding of ASD evolves, so too must the tools we use to diagnose and manage it. This project acts as a stepping stone towards this allowing user to gain a better understanding and self-awareness to help get professional help.

Acknowledgements

My deepest gratitude is extended to Professor Owen Johnson, my project supervisor, whose guidance has been really helpful throughout the project. His consistent help and insightful feedback were important for the project's success. His mentorship has not only contributed significantly to this work but has also laid a foundation for my professional and academic journey ahead.

I would also like to express my profound thanks to Professor Vania Dimitrova, my assessor, whose input was indispensable in the critical phases of development and refinement. Her observations and targeted feedback during the assessor meeting was really helpful in steering the project goals.

The group phase of this project is all thanks to the dedication and teamwork of my peers—Natalie Leung, Marilena Manoli, Chien-Wei Tung, and Archie Adams. Each brought a unique set of skills and perspectives.

Finally, a special note of appreciation is reserved for my family and friends. Their unwavering belief in me and my project progress and their moral support have helped me a lot mentally and emotionally during this project.

Together, each individual mentioned here has played a very important role in this project development. Their collective influence has not only shaped the trajectory of this project but has also left an indelible mark on my personal and professional growth.

Table of Contents

Chapter 1
Introduction and Background Research

1.1 Introduction

Diagnosing conditions like Autism Spectrum Disorder (ASD) is especially challenging because it lacks clear clinical indicators. The disorder's varied symptoms affect social interactions, communication, and behaviour[1]. These require an accurate diagnostic method [2]. There is also a noticeable difference in the level of services offered to children compared to adults with ASD [3]. For adults, the services are still developing and are marked by a lack of research, limited comprehensive treatment studies, and scarcity of new therapeutic options [3]. It's very important to understand how ASD symptoms change over time and differ between genders to improve treatment effectiveness and overall healthcare services.

Based on the data from 2018 [90], it is evident that women are diagnosed with Autism Spectrum Disorder (ASD) less frequently than men. The statistics show that only 22.51% of those diagnosed with ASD were female while 77.49% of those diagnosed with Autism Spectrum Disorder (ASD) were male, underscoring a significant gender disparity in diagnosis rates. This underdiagnosis of women may be since females often display different or subtler symptoms compared to males, which can be overlooked by traditional diagnostic methods [90]. Furthermore, as the demand for medical health services continues to grow, it becomes increasingly challenging to provide timely and effective diagnostics and treatments for all individuals. This is particularly true for adults with ASD, where diagnostic and therapeutic resources are already stretched thin [91]. To address these challenges, there is a growing reliance on artificial intelligence (AI) and machine learning technologies.

Artificial intelligence and machine learning are at the forefront of creating diagnostic tools for ASD [3]. These promise quicker diagnostics and greater independence in the diagnostic process, overcoming traditional limitations [3]. Moreover, interactive apps that mimic social situations are increasingly useful, both for diagnosis and as therapeutic tools, helping individuals with ASD in managing social complexities. In particular, this project focuses on developing improved diagnostic tools specifically for adults, with an emphasis on women. Women are often underdiagnosed with ASD due to their ability to mask symptoms, a feature that traditional diagnostic methods frequently overlook [89].

Additionally, the design of user interfaces (UI) for these tools is crucial because it greatly affects how users interact with and adopt these technologies [4]. An intuitive and well-crafted UI can enhance engagement and the effectiveness of the tool, whereas a poorly designed UI might discourage its use, no matter how beneficial the technology is [4]. By focusing on developing precise, user-centred technologies, I aim to cultivate a healthcare application that not only identifies ASD effectively but also supports and honours the experiences of adults living with the condition.

1.2 Project Aim:

The project was done over two separate stages. This journey required the integration of each team member's individual work into a single platform of the healthcare application framework. My specific contribution was to make a model prototype capable of identifying and tracking signs of autism.

1.2.1 Phase 1 Group work

During the first phase of the project, our group, which included myself and colleagues Natalie Leung, Marilena Manoli, Chien-Wei Tung, and Archie Adams, dedicated our efforts to familiarizing ourselves with the core architectural framework initially done by the previous year's student cohort, guided by the expertise of Owen Johnson. We added multiple features to the existing architecture, focusing on optimizing its structure by removing redundant libraries, enhancing its existing features, and other such improvements, all intended to improve the application’s overall functionality. Each member of the group recorded their specific enhancements to the platform using GitHub issues and Wiki.

1.2.2 Phase 2 Individual Project

During the second phase, my focus shifted towards conducting independent research and development aimed at creating a prototype AI model specifically designed for the detection of Autism Spectrum Disorder. This phase included researching different symptoms found in individuals having ASD and looking for datasets which can be used for the purpose of the project model training depending on the characteristics identified. Ultimately, the objective was to integrate the trained model into the health application (LARKS) mentioned earlier.

1.3 Project Objectives

The project was divided into 2 phases, each with its specific set of objectives to ensure structured progress.

1.3.1 Phase 1 – Group work

Prior to a deep understanding of the existing platform’s capabilities, we set out initial objectives to guide the preliminary phase of the project. These objectives were crucial in structuring the team's efforts and resources effectively and are marked by [x] to denote their completion. The extent to which these objectives were achieved, as well as the implications of their success or failure, are analysed in Chapter 3.

- Review Previous Year Platform [x]
- Identify Critical Feature Requirements for the platform to update on [x]
- Testing [x]
- Deploy Application [x]
- Load Testing [x]

1.3.2 Phase 2 – Individual Project

The second phase of the project was an individualized research and development journey, with a clear focus on addressing the challenges of detecting autism in adults.

- Research - Investigate and Identify Methodologies [x]
- Design and Calibrate Machine Learning Model – Application Feature 1 [x]
- Develop Interactive Social Engagement Tool using NLP – Application Feature 2 [x]
- Design an Eye Movement tracking Camera Detection Tool Using Machine Learning[x]
- Integrate Questionnaires [x]
- Provide Comprehensive User Feedback [x]
- Small Chatbot component [x]

1.4 Ethical Concerns

Data Privacy and Security - To protect sensitive information, all private data, like passwords, undergoes encryption processes to ensure its confidentiality. Furthermore, previous year students integrated Cross-Origin Resource Sharing (CORS) policies as an additional layer of security. CORS is a web security feature that allows us to restrict resources on a web page to be requested from another domain outside the domain from which the first resource was served [43]. By implementing CORS policies, we significantly reduce the risk of cross-site scripting attacks, ensuring that only authorized web applications have the ability to interact with our resources [43]. The application's backend hosting was done on the AWS Free Tier which provides us with a reliable and secure cloud computing environment allowing intense security.

Informed Consent - Clear and concise consent forms will be developed, explaining how data will be collected, used, and stored. Participants will be informed of their rights to withdraw consent at any time, and the process for doing so will be straightforward and accessible.

Ethical Sourcing of Data - Data will be sourced ethically, ensuring that it is either generated synthetically or obtained through channels where explicit consent for use has been given. Public domain sources like Kaggle will be used responsibly, and additional consent will be obtained if there is any ambiguity regarding the use of datasets.

Project Publication and Transparency - While the project will not be published in the public domain due to its nature as an undergraduate project, all project reports and documentation will include a disclaimer clearly stating its academic purpose.

Limitation Disclaimer - Disclaimers will be prominently displayed within the application and in any associated documentation to inform users that the tool is for educational purposes and not a substitute for professional medical diagnosis or treatment.

1.5 Risk Mitigation

Agile Sprints for Group Collaboration: We used agile sprints to ensure team alignment and timely delivery of individual contributions, making sure to have efficient workflow [66].

Unforeseen Technology Requirements: Conducting comprehensive preliminary research will help us anticipate and integrate necessary technological changes, minimizing disruptions.

Errors in Code, Breaking Application: Implementing version control with branching and merging strategies will manage coding errors effectively.

Unforeseen Functional Limitations Caused by Bad Deployment Architecture: Early planning of the application's architecture, along with maintaining a list of backup technologies, will address potential functional limitations.

Model Has Poor Results When used in the Real World: To prevent overfitting, used data augmentation techniques to enhance the model's real-world applicability.

Consistent Application Regression Testing: Regression Continuous testing, including unit, integration, and other tests, will be conducted to ensure the application's performance and reliability throughout its development cycle [66].

Further Risk Mitigation Table can be viewed in the Appendix C.3 [Risk Mitigation Table].

1.6 Project Deliverables

Upon successful completion, the project would yield the following key deliverables:

1. A comprehensive final report.

2. Access to a GitLab repository containing the designed application platform.

3. A demonstrative video showcasing the application features and functions.

1.7 User Story [66]

Sophia, curious about the possibility of being on the autism spectrum, decides to use the Autism Detector platform to gain insights. Once she sets up her account, she starts using the platform's tools like a game that helps her understand how she interacts with others and a feature that keeps track of her notes. These tools help her see areas she might want to look into more. Encouraged by what she learns, Sophia tries more advanced features of the app. She uses an eye-movement tracker that studies how she looks at things during social interactions, which is like practising real-life social situations. She also keeps using the note tracker to write down her daily thoughts and experiences, which shows patterns that might be signs of Autism Spectrum Disorder (ASD). After using these tools, Sophia moves on to filling out specialized questionnaires that the Autism Detector offers. These are designed to check for traits typically seen in adults with ASD. The results from these questionnaires confirm what she learned from the other tools and give her a clearer understanding of herself and how she might fit on the autism spectrum. With all this information and insights from the tools and questionnaires, Sophia feels ready to talk to a healthcare professional about her findings.

1.8 ASD Spectrum

Autism is one of five disorders categorized under Autism Spectrum Disorder (ASD) by the American Psychiatric Association (2000), which also includes Asperger syndrome, childhood disintegrative disorder, PDD-NOS, and Rett syndrome [8]. It appears primarily in the first three years of life, characterized by challenges in social interaction, communication skills (verbal and non-verbal), and repetitive behaviours [6][9]. The symptoms and abilities of individuals with autism vary significantly [10]. Recent research shows that autism traits are distributed on a spectrum across the population, indicating that features associated with ASD can vary widely among individuals, whether or not they are diagnosed with ASD [11].

1.8.1 Why choose adults over children?

Focusing on adults in this project addresses significant gaps in autism diagnosis, thus highlighting the lack of specialized diagnostic tools for this [13]. The concept of a 'lost generation' of adults, whose autism was unrecognized in childhood and only identified later in life, is well-documented by Lai & Baron-Cohen in 2015 [14]. Additionally, many autistic adults develop compensatory behaviours known as 'camouflaging,' which can mask the presence of autism traits and complicate the diagnostic process [15]. These individuals often encounter systemic healthcare barriers, including interactions with multiple healthcare professionals and prolonged wait times for diagnosis [16][17].

1.8.2 Use of AI

The incorporation of artificial intelligence (AI) in supporting individuals with Autism Spectrum Disorder (ASD) has been an area of growing interest and development. Putnam and Chong, in their 2008 study, discuss the positive reception of software and technologies among users with ASD, indicating a favourable user experience and potential for such tools to offer meaningful support [34]. Machine learning (ML) is a widely used area in AI research [55]. ML excels in pulling out important details and forming advanced models that often do better than humans in managing big data sets [56]. Such capabilities of ML can improve our grasp of Autism Spectrum Disorder (ASD) and potentially lead to improved methods for its screening and diagnosis [57].

Historically, assistive technologies and intervention programs have primarily targeted younger demographics—focusing on supporting young children, school-aged children, and adolescents [35]. These resources have been used towards enhancing social skills within educational environments or assisting with daily routines, as noted by Tomczak [35], and further evidenced by interventions described by Alabbas & Miller [36], Caldwell [37], and Fage et al. [38][39].

However, this project recognizes a significant gap in the available technologies aimed at adults with ASD. It seeks to bridge this gap by integrating technological advancements, particularly AI, with the unique needs of autistic adults—a demographic that has not traditionally been the primary focus of such interventions. By doing so, the project aims to extend the benefits of technological support to include adults with ASD, providing them with tools and resources that are tailored to their specific requirements and challenges. This shift in focus represents a crucial expansion of the scope of assistive technology and programs within the field of ASD support.

1.8.3 Existing Solutions and Techniques

Today, various tools help diagnose autism spectrum disorder in adults. One such example is highlighted by Song et al. (2019), who reviewed the literature on using artificial intelligence (AI) to enhance the diagnosis process. They found multiple studies that apply different AI techniques to current assessment methods to create models for identifying individuals with ASD, detailed in tables 1 and 2 of their paper, available in Appendix B.2 [57]. However, developing these sophisticated machine learning models requires significant time, resources, and manpower, making them impractical for this particular project. Instead, this project will use these models as a reference point to guide research on development tools. Most general tools for ASD screening are simple questionnaires that follow a standard format. This project aims to add features that use these questionnaires (Autism Spectrum Quotient, AQ10, CAT-Q, RAADS-R) as a supplemental tool for final analysis and to guide users towards seeking professional advice. These enhancements will be discussed in more detail in the upcoming Chapters 2, 3, and 4.

1.8.4 Autism Spectrum Quotient [20]

The Autism-Spectrum Quotient (AQ), developed as a measure for autistic traits in adults, especially those with an average or above IQ, plays a crucial role in autism research and clinical applications [12]. Recognizing that at least half of the individuals on the autism spectrum display a wide range of experiences and traits, the AQ is designed to reflect this diversity effectively [18]. It presents respondents with 50 statements, asking them to express their level of agreement using a four-point scale ranging from 'definitely agree' to 'definitely disagree' [11]. This method captures a detailed profile of autistic traits, using a binary scoring system where agreeing with a statement scores +1, leading to a maximum possible score of 50 [20]. The scoring system, alongside a balanced design to prevent response biases, ensures the AQ's reliability [11]. However, it's tailored for individuals who can fully understand the questions, suggesting alternative assessments for those with significant cognitive or language barriers. The AQ's meticulous construction and validation enhance its utility in understanding and assessing autistic traits among adults [11].

1.8.5 AQ-10 [40]

The AQ-10, an abbreviated version of the Autism-Spectrum Quotient (AQ), represents a significant advancement in ASD screening. Developed through a study by Allison et al. in 2012 [19], it involved a sample of adults, both with ASD and control groups, refined through calibration and validation phases. Further validation through ROC (receiver operating characteristic curve) analysis by Booth, T., Murray, A.L., McKenzie, and colleagues [21] confirmed the AQ-10's high sensitivity, specificity, and positive predictive value, establishing its efficacy as a "rapid screener." This highlighted the need for further validation with independently verified ASD diagnoses to enhance its reliability [21]. A suggested cut-off score was also proposed to optimize its utility in scenarios constrained by time [21]. Following these findings and recommendations from the National Institute for Health and Clinical Excellence in 2012 [22], my project implemented both the full AQ and the AQ-10 to provide comprehensive insights while accommodating the practical limitations of frontline professionals. This dual approach aims to offer a thorough assessment of potential ASD-related traits, enhancing the diagnostic process.

1.8.6 CAT-Q [41]

The CAT-Q, or Camouflaging Autistic Traits Questionnaire, addresses a critical aspect of many autistic individuals' experiences: social camouflaging. Defined as the strategies used by autistic people to mask their autism traits in social contexts, this concept was notably explored in a pivotal study by Lai et al. in 2011 [23, 24]. This research highlighted that social camouflaging is particularly prevalent among autistic women and girls, who may not receive a timely diagnosis due to their ability to 'mask' or adopt non-autistic personas to navigate social norms (Bargiela et al. 2016; Tierney et al. 2016) [25, 26], and adjust their behaviours to align with peer expectations, reducing apparent social and communicative difficulties (Dean et al. 2017) [27]. The research also shows the nature of autism in females, emphasizing how the condition's presentation and societal perception differ, often leading to a lack of recognition (Mandy et al., 2012; Kopp and Gillberg, 2011) [30, 31]. Gender differences in social behaviour, such as those studied by Dean, Harwood, and Kasari in 2017, reveal that societal gender biases may enable girls with autism to more effectively mask their symptoms [27]. These societal biases also influence how autistic behaviours are perceived and tolerated within various social settings (Lai and Baron-Cohen, 2015) [14]. Research further indicates that while camouflaging can serve as a coping strategy, it often results in increased stress and a higher vulnerability to mental health issues such as depression [28, 29]. It also complicates the accessibility to clinical support the services, as the camouflaging behaviours keep the underlying issues unrecognized (Cage et al. 2017; Head et al. 2014) [28, 29].

1.8.7 RAADS-R [42]

The RAADS-R, or Ritvo Autism Asperger Diagnostic Scale-Revised, developed by Ritvo et al. in 2008 [32, 33], is a refined tool for diagnosing Autism Spectrum Disorder (ASD) in adults. It builds on the original RAADS and complements the Autism-Spectrum Quotient (AQ) by Baron-Cohen et al. [12]. The RAADS-R includes 80 questions, adding a new area—circumscribed interests—and improving question clarity [32]. Designed for adults with average or higher intelligence, the RAADS-R features a dual-question structure: 64 symptom-based queries scored from 3 ("true now and when I was young") to 0 ("never true"), and 16 normative behaviour questions, scored inversely [32]. It's divided into four subscales based on Diagnostic and Statistical Manual of Mental Disorders, DSM-IV-TR criteria: Social Relatedness, Circumscribed Interests, Language, and Sensory-Motor [32]. Validated in 2011, the RAADS-R demonstrated high diagnostic precision, distinguishing effectively between ASD individuals, those with other DSM-IV-TR diagnoses, and those without diagnoses [32]. It showed excellent diagnostic sensitivity, specificity, and test-retest reliability, with an overall accuracy of 98.5% [32].

1.9 Amazon Web Services (AWS)

One of the projects objectives included research and deployment of the application. We have selected AWS Free Tier for deploying our project's backend due to its robust features and enhanced security, as detailed in Chapter 3. The National Institute of Standards and Technology (NIST) defines cloud computing as a model that enables ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources, such as networks, servers, storage, and applications. These resources can be rapidly provisioned and released with minimal management effort [54]. Amazon Web Services, being a form of cloud computing, offers a range of services categorized into Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). It provides essential computing storage and network capabilities, platforms for deploying custom applications, and integrated infrastructure and software for running office applications [53].

1.10 Old Architecture Evaluation

The first phase of the project focused on researching application and their frameworks done by the previous year's (2022 cohort) students [70, 71]. Upon further researching, the group decided to improve on one of the previous year's application architecture frameworks shown in Fig B.1 Appendix B [65].

The backend is a Flask application running on an Amazon EC2 instance, which serves as the application server. It includes individual applications with their corresponding databases for user-specific data. A login database within the Flask backend manages authentication, leveraging Flask-Bcrypt for secure password storage. The frontend is built as a React application, providing a user interface. It retrieves static content from an Amazon S3 bucket, which stores the React files. The React Camera Module enables direct interaction with the device's camera, an essential feature for a medical application. Communication between the client's device and the AWS-hosted services is secured using HTTPS. Amazon CloudFront, a content delivery network (CDN), sits between the user's device and the AWS services to expedite the delivery of content by caching it at edge locations.

Last year's architecture showed some key advantages and disadvantages. The use of a React frontend was beneficial as it optimized the user experience through fewer page reloads, typical of Single Page Applications, resulting in a smoother and more responsive application. Performance also got a boost from Amazon CloudFront, which helped cut down delays and quicken content delivery with its caching abilities. Scalability was another strong point. The EC2 instance could scale to meet demand, and CloudFront's ability to distribute the load across its network meant the application could manage more users without a hitch. The architecture's division of the frontend into shared and individual components promoted code reuse and made updates and maintenance more straightforward.

On the other side, the architecture had some drawbacks. Fault isolation was limited; issues in a single component could potentially affect through the entire application, creating a single point of failure. Deployment flexibility was also a concern; any change necessitated a full application redeploy, leading to longer downtimes. Finally, scaling was often a complex task. Usually, the only way was to size up the EC2 instances, which can be expensive and has practical limitations, making it a less-than-ideal solution when resource demands increased.

1.11 Individual Application Technologies

The application for the Autism Detector followed the same framework as the shared platform did – Python Flask and React with the databases using SQLite and SQLAlchemy [73]. It also uses an RNN structure for some of its features mentioned in Chapter 3 and 4. Below is a detailed explanation of the RNN Architecture and why it was used.

1.11.1 RNN Architecture

Neural network language models (NNLMs) traditionally have just one hidden layer, which can only handle basic input features. However, a deep neural network (DNN) with multiple layers is better equipped to deal with more complex features [44]. Recurrent neural networks (RNNs) differ because they can remember recent inputs using feedback connections, unlike long-term memories that are stored in weights that change slowly [45]. Early research on the Long Short-Term Memory (LSTM) model, a specific type of RNN, demonstrated that insights from simpler RNNs are also applicable to LSTMs [46].

RNNs, however, come with certain limitations [47]. Typically, an RNN's state, whether it's a simple RNN or an LSTM, depends only on past and current events. The state at any time is influenced by events up to that moment [48]. But in some cases, accurate predictions require knowledge of past, present, and future events [48]. A bidirectional recurrent neural network (BRNN) addresses this by using information from both past and future within a specific timeframe, dividing the state neurons into two sections: one for forward time direction (forward states) and one for backward time direction (backward states) [47]. In the note tracking feature of the individual application Autism Detector, an RNN equipped with LSTM units is utilized, as thoroughly explained in Chapter 3. This approach leverages the RNN’s capability to sequentially process information, making it well-suited for monitoring the progression of notes in the application.

1.11.2 Using Natural Language Processing

Natural language processing (NLP) is a theory-motivated range of computational techniques for the automatic analysis and representation of human language [50]. One of the features of this project uses NLP. This has been further mentioned in Chapter 3 and 4. Below is a detailed summary about the research done on NLP before its implementation.

1.11.2.1 Semantics-based NLP

Semantics-based NLP focuses on the intrinsic meaning associated with natural language text [50]. Rather than simply processing documents at syntax-level, semantics-based approaches rely on implicit denotative features associated with natural language text, hence stepping away from the blind usage of keywords and word co-occurrence count [50]. A guide to semantic-based NLP was used for its implementation in this project [51].

1.11.2.2 Sentiment Analysis Using Transformer-Based Models

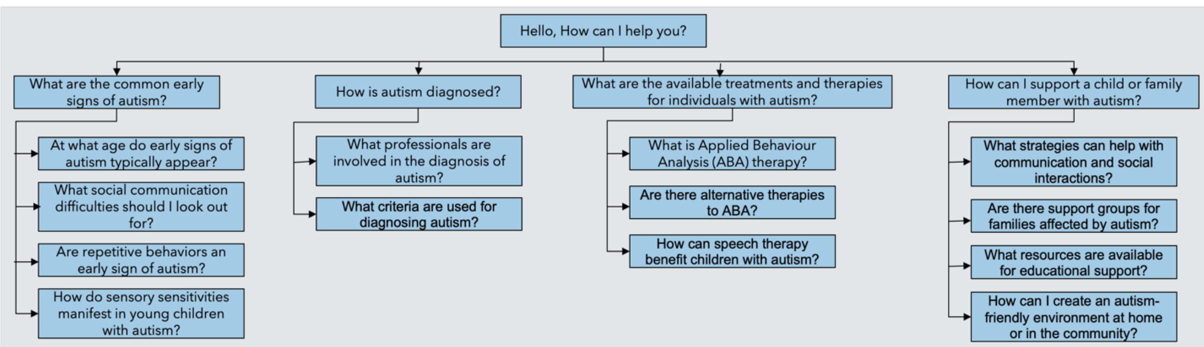
Transformer-based models are advanced tools for processing language. They use a special setup that includes an Encoder-Decoder structure and self-attention mechanisms, which help them perform very well [51]. Although it is possible to build these models from scratch, this process is usually complicated and takes a lot of time [52]. Instead, we can use models that are already made and available on Hugging Face, a platform that offers many models for language tasks [52]. This project uses ready-made transformer models from Hugging Face, as detailed in Chapter 3. A guide was also followed to add this feature to the project [51].

1.12 Stakeholders

NHS/Doctors - For the NHS, the model represents an opportunity to enhance diagnostic processes, potentially reducing the time and resources required for autism spectrum disorder (ASD) assessment. By integrating this model into clinical workflows, the NHS could provide more timely and accurate diagnoses, improving patient outcomes and resource allocation.

Users (age 18-30) - For the users, particularly those aged 18-30, the model offers a screening tool that could help with the detection and tracking of ASD. The user can then use this application as a stepping tool for self-assessment and use the informed feedback as a report they can show to their NHS doctors.

1.13 Fig 1: Simple Chatbot Decision Tree – used to show information to user



Chapter 2
Design and Implementation (Methods)

2.1 Project Workflow

Phase 1 of the project was done in the form of weekly sprints in semester 1 which were defined and tracked using GitHub wiki. Each week had set goals for the group. A scrum master (Marilena) was appointed to help oversee the progress of the project. The scrum master along with the other members of the group got together every week for a meeting to discuss the progress and the updates and the future work for the project.

Phase 2 of the project followed a waterfall method [66] which allowed a set timeline for the project and its objectives. It also allowed each member of the group to work on their individual applications without further disturbance or group requirements unless the common framework of the application was affected.

2.1.1 Gantt Chart [92, 93]

The project workflow was documented on GitHub wiki using issues and can be seen using the below Gantt Charts [92, 93].

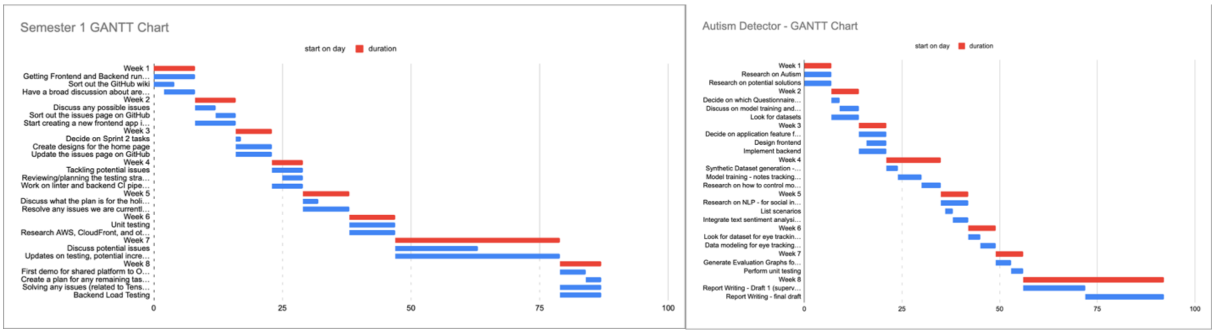


Fig 2: Thumbnail for Gantt Charts, larger image in Appendix C.4

2.2 Improved Architecture

The backend of the project was deployed using AWS EC2 instance. Following the GitHub wiki documentation from the previous year and error handing due to changes in device compatibility and other issues, the backend was hosted on the EC2 instance link - Public IPv4 address: http://13.42.9.210/. The frontend connection to the backend was done through CloudFront with the following link: Distribution domain name: https://d28uu5wwh8m9ry.cloudfront.net. The frontend was deployed on GitHub, by a team member Archie Adams on the link: https://archie-adams.github.io/larks. We designed the application architecture to optimize both function and security, employing AWS and GitHub for efficient deployment. I deployed the backend of the application in an AWS EC2 instance operating within a default VPC, creating an isolated network that enhances security. This instance runs the Flask backend. It must be noted that due to the storage limitations and nginx issues, further discussed in Chapter 4 and 5, the application wasn't completely deployed as the backend and frontend isn't properly connected. If the backend Public IPv4 address is accessed, it will show a 'Home page' as a means to check its deployment and testing.

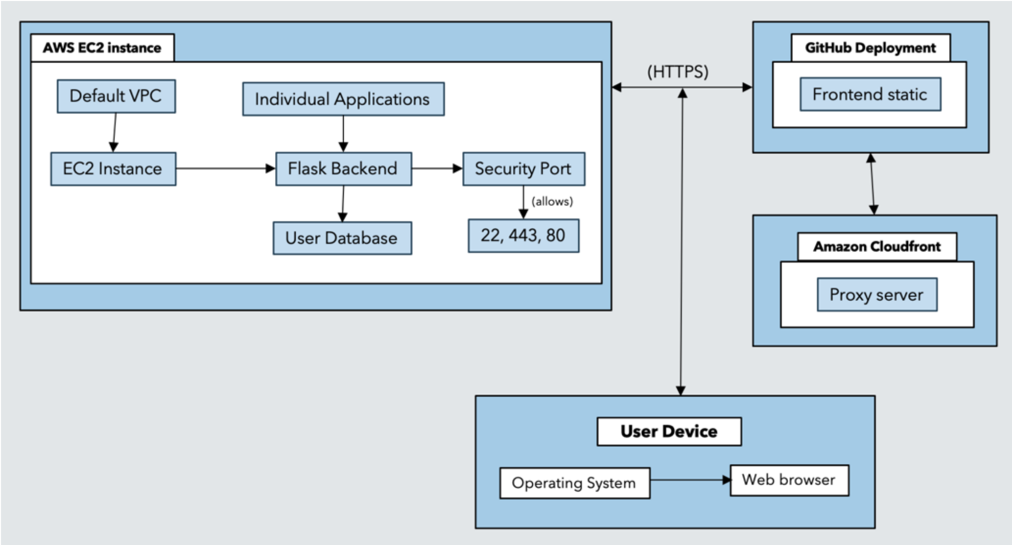


Fig 3: Deployment Architecture Diagram

The heart of user data management lies in the User Database, which is linked to the Flask backend, within the same VPC or on the EC2 instance itself, ensuring swift and secure access to user information. Security port range is placed, and specific ports are open to facilitate secure communication; port 22 for SSH, port 443 for HTTPS, and port 80 for HTTP. However, the system is configured to redirect all HTTP traffic to HTTPS, ensuring all data transferred is encrypted.

For the user-facing side of the application, the frontend static files are stored in a GitHub repository, allowing for version control and collaborative development. The frontend was deployed by a group member, Archie. These static files include all elements of the user interface such as HTML, CSS, and JavaScript, and are consistently served to each user without dynamic changes.

To deliver these files to the end-user, I employed Amazon CloudFront, a CDN (Content Delivery Network) service that caches these static files across global servers. A CDN is an essential element in most contemporary web applications, enhancing content delivery by duplicating frequently accessed files (static content) across a globally dispersed network of caching servers [94]. This not only reduces latency but also improves the performance of the web application. CloudFront further serves as a proxy server, handling requests by routing them appropriately between the user and the backend on the EC2 instance.

Unfortunately, I was unable to get it to work accurately as mentioned earlier and the link is broken between the backend and the frontend due to reasons including separate deployment sites and nginx direction. This part of the project requires further work. On proper development of the framework, the user will interact with the application through their device, equipped with a web browser and an operating system. The web browser communicates with the CloudFront service over an HTTPS connection, guaranteeing that all information remains secure and encrypted.

Finally, the architecture emphasizes security with the HTTPS protocol encrypting all communication between the user device and the frontend. This approach ensures that data integrity and security are maintained at all times, providing a seamless and secure experience for the user. This setup effectively connects the backend and frontend components of the application, utilizing AWS for hosting, GitHub for code deployment, and Amazon CloudFront for efficient content delivery and enhanced security, showing a modern, cloud-based application deployment strategy.

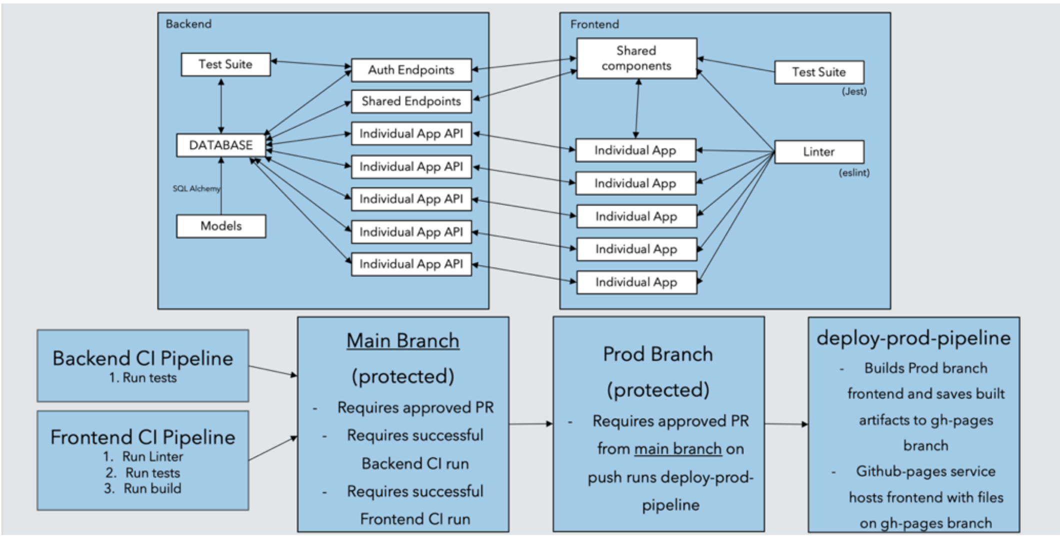


Fig 4: Component Architecture diagram (made by Archie)

We've updated the backend part of our web application to make it better at handling data and keeping user information safe. We used SQLAlchemy in our Flask-based system to organize our database models better. This makes managing data and checking it for errors easier. We've also made some big improvements based on problems we saw before. One such improvement was making sure passwords are more secure. One of the group members, improved the password encryption by building on what students did last year, using Flask-Bcrypt to make sure user passwords are kept safe.

We also made sure to use HTTPS throughout our system to protect data when it is sent over the internet and stop potential spying or attacks. Archie played a key role in setting up new features - Continuous Integration/Continuous Deployment (CI/CD) pipelines for both the backend and frontend parts of our app. This lets us update parts of the app separately and more often without much downtime, which means we can bring new features and fixes to users faster along with checks for the application before the updated version is merged into the main application repository. These checks included Jest tests, ESLint and setting up Node.js for frontend. For the backend Python 3.8 requirement checks and other tests were done as the application was built on this version of python.

In managing our databases, we used SQLAlchemy, which makes it easier to switch between different database systems and improves how we manage them. This helped us not rely too much on specific services like those from AWS. We continued to focus on testing and making sure our app has high-quality code with few bugs by using thorough testing systems and code checkers for both the backend and frontend.

We've also made sure to define specific areas in our app for user login and security checks, which helps us better monitor and protect these processes. Automating the deployment process through a specific production pipeline reduces mistakes and lets developers focus on more important tasks. Using protected branches for our main and production code means only the best-tested code gets pushed on, which cuts down on problems from unstable updates.

2.2.1 Login tests

During the review of the previous year's system, our group noticed that there was a gap in testing as it is an important aspect of any project [67]. To address this, we decided to carry out comprehensive testing, with my personal focus being on the SignIn functionality. I made sure the login form displayed all the necessary elements correctly. I also checked to make sure that if someone tried to log in with empty fields, the system would show an error message. It was important to confirm that the system would accept the form with the right credentials and that it would take the user to the home page after a successful login. The link to sign up for new users was another area I tested; I made sure it led to the correct page. I checked how the system responded to server errors, ensuring that it would show the right message if something went wrong on the server's end. This included handling incorrect login details, problems when there was no internet connection, and delays due to the server taking too long to respond. I also tested for the correct response to unexpected server errors. Lastly, I made sure that after a successful login, the system would correctly set the authentication token, which is crucial for users to stay logged in and secure.

2.3 Individual Model Architecture

The application is based on a shared connected database that stores the user data and results. The backend manages the database connection using SQLAlchemy. The application has two major features using model training (Notes tracking and Eye tracking) and one feature (social interaction game) focusing on using Natural Language Processing (NLP). The model training and NLP for the application are explained in detail in Chapter 3. The below figure shows the Architecture diagram for my Individual Application, Autism Detector.

For model training, my architecture incorporates two distinct processes: one for textual data and another for eye-tracking data. The textual data process involves generating a dataset, which is then pre-processed using NLTK to clean and prepare the text. The data is tokenized using Keras, and a machine learning model is defined and trained using a combination of embedding layers, bidirectional LSTM layers, and dense layers. This model is rigorously evaluated using k-fold cross-validation before being saved with Joblib for later use. Similarly, the eye-tracking data process involves loading and pre-processing data, merging it with metadata, and engineering features to train a binary classification model. This model is optimized with GridSearchCV and also validated through cross-validation before being saved. These processes uses tools such as SQLAlchemy for database interactions, and Python libraries like Keras and Joblib for machine learning tasks, providing a comprehensive system for handling complex datasets and training models to deliver personalized user experiences or detailed analytics. These have been explained in detail in Chapter 3.

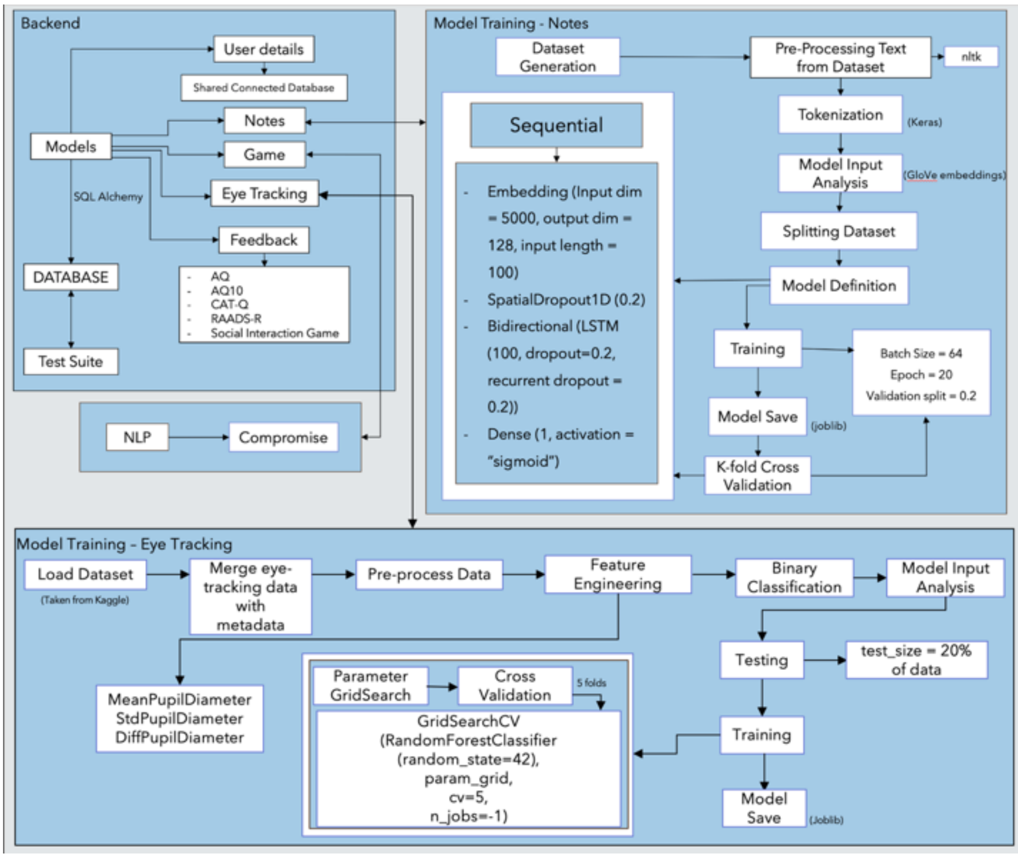


Fig 5: Individual Application Architecture Diagram

The database schema is shown in Appendix C.5 with all the tables used for Autism Detector. The information stored in the database such that each user's unique identity and security is maintained through email and password fields, allowing for secure and personalized access to the app's features. Detailed personal information including demographic details, educational background, and additional conditions like sensory sensitivities are captured to tailor the user experience and enable the analysis of how different factors influence autism spectrum characteristics. The GameFeedback table collects data from interactive game scenarios, which helps in evaluating responses that are indicative of ASD traits. Observational notes and time-stamped entries provide a record of user interactions and changes over time, crucial for tracking the effectiveness of interventions. Finally, diagnostic feedback from the questionnaires is used to enhance the accuracy of autism detection and providing valuable insights into individual and collective ASD profiles.

2.4 User Interface Development

The design and implementation of a user interface (UI) are critical to this project, serving as the main way users interact with the tool. A well-designed UI enhances user experience, encouraging adoption and favourability, while a poor one can lead to rejection [4]. I designed the shared platform home page as well as my individual application home page using the guidelines from the UK Department of Health [5], initially for documents for those with learning disabilities, to improve UI accessibility. These include using text with visual aids for clarity, positioning images on the left of the associated text, employing large, legible fonts like Arial of at least 14 points, and limiting sentence and document length for easier comprehension. Additional UI design principles from Pavlov, N., (2014) [4] include using contrasting colours for text and backgrounds. I avoided bright colours that overwhelm users.
Throughout development, two UI designs were created for my individual application and refined through feedback from neurotypical individuals and those with ASD, ages 18-30 using a Google form with anonymous feedback [64]. 55.3% of the 85 participants were uncertain about their neurotype, reflecting a significant awareness gap and suggesting a need for enhanced educational and diagnostic opportunities. Using the feedback collected from the designed Google form, Design 2 was selected. The application's user interface was changed according to the feedback to reflect the preferences and requirements expressed by the participants of the google form. Below images show the sigup and login pages as well as the home page once the user signs up. The home page was used as the shared platform for each individual application and Autism Detector, is the seventh in the home page.

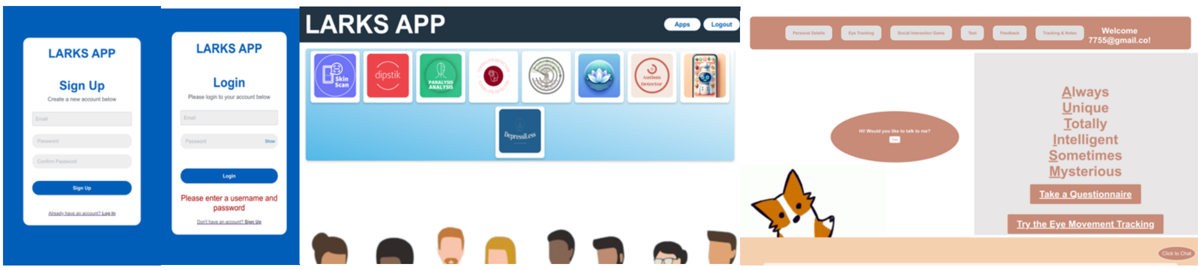


Fig 6: Application Home page. Please look at Appendix C.6 for larger image

Chapter 3
Methods – Individual Project

3.1 Individual Application Implementation

After the shared platform was set up and running, each member of the team worked on individual applications during the second semester of the project. I developed various components for the Autism Detector application. Firstly, I defined and constructed a database using SQLite, tailored to meet the specific needs of the application, following a carefully designed database schema which can be looked at in Appendix C.5. Next, I developed the frontend, creating user interfaces that allowed users to input data and complete tests, providing them with feedback based on the results. Other components included an NLP social interaction scenario game, a model for tracking notes, another model for tracking eye movements, a simple chatbot for user interaction using a decision tree shown earlier in Chapter 1.13.

3.2 NLP (Social Interaction Game)

The NLP (Natural Language Processing) part of the code involves using the compromise library, to process user responses in a game designed to enhance social interaction skills. This library analyses user responses to predefined scenarios where each scenario requires a response that demonstrates their understanding or approach to a given social situation. The function nlp(response).terms().out('array') is used to check if the responses include correct or inappropriate keywords linked to each scenario.

Additionally, it uses Hugging Face's components for more sophisticated text processing tasks. It employs models like distilbert-base-uncased-finetuned-sst-2-english and their tokenizer with AutoTokenizer.from_pretrained and AutoModelForSequenceClassification.from_pretrained for loading and handling text tokenization and model tasks. Sentiment analysis and text classification pipelines are also created using these loaded models and tokenizer. The pipeline function from Hugging Face offers a straightforward API for these processes.

When a user submits a response, the application assesses this input to determine if it aligns with the expected or appropriate keywords. The game generates feedback based on the presence (or absence) of correct keywords or their synonyms in the user’s response, including calculating a score for how many correct terms were used. It also analyses the sentiment of the response through an Axios request to an API endpoint (analyzeSentiment function in the backend). It sets two pipelines: one for sentiment analysis and another for intent classification. These pipelines utilize the loaded models to process text inputs, allowing for real-time evaluation of both the sentiment and the intent behind a user's text. Using the NLP compromise library alongside Hugging Face's sophisticated tools is crucial for the game to provide immediate, relevant feedback on the user's ability to navigate social scenarios effectively, which is vital for learning and improving social interaction skills.

3.3 Machine Learning Model Training (Notes Tracking)

For this feature of the application, due to lack of dataset, the dataset used for this model training was synthetically generated - a collection of synthetic notes stored in JSON format. The first objective was to transform raw textual data into a structured format suitable for feeding into a machine learning model, ensuring data quality and preparing the input for high model performance. This dataset, was loaded into a Pandas DataFrame which then went through detailed data cleaning and normalization including lowercasing, removing non-word characters, stop words removal and lemmatization, spell correction, language detection, and sentiment analysis. Through these detailed pre-processing steps, the dataset was transformed from raw, unstructured text into clean data ready for the application of machine learning algorithms. Ensuring data integrity is an important step in the pre-processing phase, as the presence of missing values, duplicate rows, and outliers can significantly affect the performance and validity of machine-learning models.

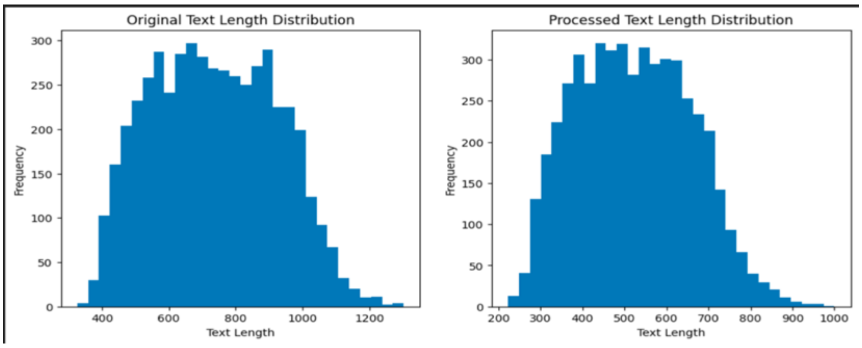


Fig 7 : Shows comparison between the distribution of text lengths before/after processing.

Another feature of the model training process included word embeddings. Word embeddings are a form of word representation that allows words with similar meaning to have a similar representation. They are a set of feature learning techniques in natural language processing (NLP) where words or phrases from the vocabulary are mapped to vectors of real numbers[95]. Embeddings capture the semantic relationships between words, so that words that are used in similar contexts are located in close proximity to one another in the embedding space. I used a pre-trained model GloVe (Global Vectors for Word Representation) [96]. It is a method that learns to represent words as numerical values by counting how often words appear together in a large amount of text. The embedding coverage check is an important step to assess how well a pre-trained model like GloVe represents the specific vocabulary used in my dataset. This step involved calculating two key metrics: the percentage of the dataset's vocabulary that is covered by GloVe and the percentage of the total text (on a token basis) that is covered.

The count of words found in the embeddings and the count of those not found (Out-Of-Vocabulary or OOV words) are tallied. Then the coverage metrics are calculated, including the percentage of the unique words in the dataset that are found in the embeddings (vocabulary coverage) and the percentage of the total word count from the dataset that is represented in the embeddings (text coverage). The loading and utilization of GloVe embeddings is important to ensuring the model has a pre-understood linguistic base to learn from. Upon evaluating these embeddings against the generated dataset, I observed that there was no pre-processed text that was not already in the GloVe model—a 100% coverage. This showed a clean dataset, with pre-processing steps perfectly aligned with the GloVe vocabulary.

In the project, I employed a Sequential model architecture for the processing of textual data. This approach is ideal for the requirements due to its simplicity and effectiveness in managing a single data stream from input to output. Starting with an Embedding layer, I transformed word indexes into fixed-size dense vectors, focusing on its ability to interpret words within a sentence for understanding language nuances. To enhance model robustness and prevent overfitting, a Spatial Dropout layer followed, which omits entire feature maps, so that the model is not dependent on any single word. The inclusion of a Bidirectional LSTM layer was important, as mentioned in Chapter 1, using its ability to learn from data sequences in both forward and reverse directions, thus grasping context more effectively than traditional unidirectional models.

$$BiLSTM(v) = LSTM(v) \oplus LSTM(reverse(v))$$

Finally, the architecture uses a Dense output layer, using a sigmoid activation function to give the probability of binary outcomes.

$$Output = activation(dot(input, kernel) + bias) \text{ [97]}$$

Here, $Output = \sigma(w^T BiLSTM(v) + bias)$

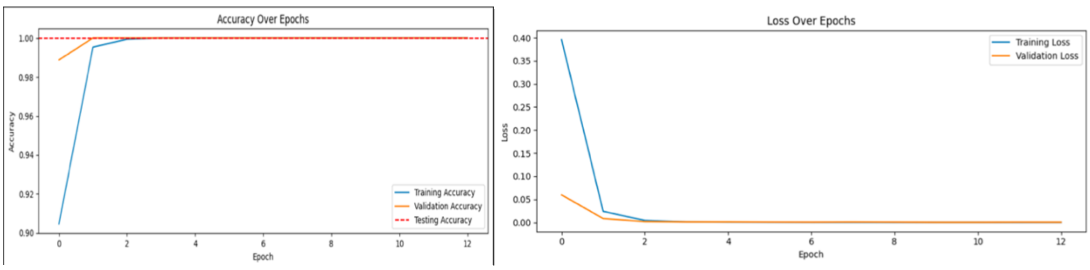


Fig 8 : Shows accuracy and loss over 20 Epochs

In the training phase of my model, the pre-processed data is used for the model training process. Through the 'fit' method, I trained the model in learning from the data over several iterations, known as epochs, while using a specific batch size for each step. To make sure it doesn't overfit, I implemented an EarlyStopping callback. This function monitors the validation loss using binary cross entropy loss function and stops training if the loss starts to decline, using the patience parameter to delay the stop for a few epochs to confirm the trend [98].

$$Loss = \frac{1}{N}\sum_{i=1}^{N} -(y_i \log(p_i) + (1 - y_i)\log(1 - p_i))$$ [98]

Where, pi is the predicted probability of class 1 and (1-pi) is of class 0 and yi is the true label This callback also restores the weights from the most successful epoch, making sure to retain the best version of the model. The validation split makes sure that there is a segment of the data reserved for assessing the model's performance, separate from the training process. This partition is important to test the model's ability to generalize to new data, beyond what it has learned during training.

3.4 Machine Learning Model Training (Eye Tracking)

In this project for the eye tracking feature, a machine learning model was developed to classify individuals as autistic or not based on eye-tracking data. The data was taken from Kaggle [72] showing eye-tracking measurements taken from a group of 25 participants, with metadata detailing participant classifications. This information was used as a single dataset, ensuring compatibility between the eye-tracking and metadata formats. Numerical transformations were conducted on key features, such as pupil diameter, to prepare the dataset for analysis. Features were engineered, including the mean, standard deviation, and difference in pupil diameters, to provide a foundation for the classification task. The 'Class' variable, denoting autism diagnosis, was binary-encoded to use in the model's learning process. A feature selection process separated the most predictive features, and a normalization step was applied to scale the feature values.

The dataset was split into training and testing subsets, with 80% of the data used for training the model. A RandomForestClassifier [68] was chosen for its efficacy in handling binary classification tasks. To optimize the model, a GridSearchCV approach was utilized, identifying the most effective hyperparameters through cross-validation. On completion of the training process, the machine learning model gave a set of parameters. The best configuration for the RandomForestClassifier was determined to be a max depth of 30, a minimum sample split of 2, and the use of 100 trees (n_estimators). This combination of parameters was identified through a GridSearchCV process that also returned a cross-validation score of 0.76, indicating a solid generalization capability on the unseen validation data.

The model was then tested and achieved a 77% accuracy, meaning it correctly predicted the class (autistic or non-autistic) 77% of the time. It showed precision values of 0.77 and 0.78

for non-autistic and autistic classes, respectively, indicating good accuracy in labelling both groups. However, the recall scores were 0.92 for non-autistic and only 0.51 for autistic classes, showing it was better at identifying non-autistic individuals. The f1-scores, which balance precision and recall, were 0.84 for non-autistic and 0.62 for autistic classes. In terms of macro and weighted averages across classes, the model scored around 0.73 and 0.76, respectively, for the f1-score. Overall, the model's performance across various metrics like f1-scores suggests it manages a decent balance in identifying classes, but there's potential to improve, especially in detecting autistic cases more reliably.

3.5 Security

Several methods used for data privacy and security were used due to the increasing need to protect sensitive information from unauthorized access, breaches, and various cyber threats. A key component of this was encrypting all sensitive data, like passwords. By encrypting this information, it becomes unreadable to anyone who doesn't have authorized access, protecting it from malicious activities. This was implemented the previous year (2022-23) and refined this year by one of the group members.

In addition to encryption, I recognize the specific challenges that arise from web-based interactions. To address these, Cross-Origin Resource Sharing (CORS) policies was implemented using previous years application code. CORS is a vital security mechanism that restricts the resources on a web page to interact only with requests from domains that have been approved. This significantly lowers the risk of security breaches, particularly from cross-site scripting (XSS) attacks, where hackers attempt to inject malicious scripts into web pages to steal data or cause other harm.

Choosing to host my application on the AWS Free Tier also plays a crucial role in maintaining a secure and reliable operation. AWS offers a robust set of security features that I utilized. These controls help me control who can access specific resources, enhancing the security framework of my application. By using these tools and policies, I can ensure my application is not only secure from threats but also operates efficiently and reliably under the protective measures provided by AWS.

Another form of security for the individual application included unit tests for each function of the application was done along with testing for components separately [66] of individual applications done the previous year. Testing the backend primarily focused on the database and its uses making sure to follow robust procedures to check all GET and POST requests implementing error handling: HTTP 400 Bad Request, HTTP 404 Not Found, HTTP 500 Internal Server Error.

Chapter 4
Results and System Evaluation

This Chapter will focus on the results of the entire project gathering the conclusive data for the Autism Detector application and going into detail about the project deliverables as well as the user feedback formatting.

4.1 Results of Testing Shared Application on Different Devices

Reachable by backend deployed link http://13.42.9.210/

The link is reachable by the following devices – Mac, Windows, Linux, iPhone and Android.

4.2 Shared Platform Load Testing - Backend

I conducted Load testing to assess the performance of our backend services under simulated conditions of increasing user traffic. The tests were executed on a MacBook Pro with the following hardware specifications: Processor: 8-Core Intel Core i9, 2.3 GHz, Memory: 16 GB, OS: macOS. The tests focused on a series of user interactions mimicking real-world usage patterns, aiming to identify potential bottlenecks and to ensure that the system can handle anticipated traffic levels without degradation of performance.

4.2.1 Test Configuration

I used Artillery, a powerful load testing toolkit, to generate traffic towards our backend deployed locally. The test consisted of two main phases:

Ramping Up: Starting from 1 user per second and gradually increasing to 5 users per second over a period of 120 seconds.

Sustained Load: Maintaining a constant load of 5 users per second for 300 seconds.

Each generated user executed a predefined set of operations including registration, login, and data retrieval, all critical operations for our service.

4.2.2 Key Metrics

HTTP Status Codes: Ensuring the majority of responses were successful (HTTP 200).

Response Times: Tracking the minimum, maximum, mean, and percentile response times.

Concurrency Metrics: Monitoring user sessions, creation rates, and failure rates.

4.2.2.1 Results Overview

Success Rates: A high success rate was observed, with a total of 5,580 successful responses indicating robust error handling and service stability.

Performance Under Load: The mean response time remained consistent even as the load increased, averaging at 433.6 ms. The system sustained its performance without significant spikes in response times, which peaked at 1,434 ms under the highest load.

System Throughput: The system handled an average request rate of 8 requests/sec, with peak rates reaching 15 requests/sec during sustained load phases.

The detailed performance load testing results can be found in Appendix E. In the detailed performance insights from the load testing, several key metrics stand out. The maximum response time observed during peak load was 1434 milliseconds, which represents the slowest response under the highest stress conditions. The average response time across all transactions remained relatively low at 433.6 milliseconds, suggesting that the system handles average loads efficiently. Furthermore, the 95th percentile response time was 699.4 milliseconds, meaning that 95% of all requests were processed faster than this time. Similarly, the 99th percentile response time stood at 742.6 milliseconds, indicating that nearly all requests were handled swiftly, thus demonstrating the system's stability even as it approached its performance limits. These metrics collectively suggest that while the backend performs well under typical conditions, there is room for optimization to reduce the impact of peak loads on response times. The load test results demonstrate that the backend system is capable of handling increased traffic with minimal response time variance, ensuring a smooth user experience during peak load conditions. However, the presence of a maximum response time of 1434 ms suggests there are potential areas for optimization. This has been discussed further in Chapter 5 Future Work.

4.3 Individual Application Model Results

4.3.1 Note Tracking

(Precision: 1.0, Recall: 0.99, F1-Score: 0.99, ROC AUC: 1.0)

Following the model's precision, it is highly likely that the model is overfitting due to the small size of the synthetically generated dataset. To reduce the chances of the model overfitting, K-fold cross-validation was done.

4.3.1.1 K-Fold Cross-Validation

To make the results more consistent, the validation error can be averaged across multiple sets of data. This process is known as cross-validation, and the most common type is called K-fold cross-validation [74]. This technique divides the dataset into 'k' subsets, or folds, of same size. In the training phase, each model iteration is exposed to 'k-1' folds as the training ground, leaving the other fold as the proving ground for validation [74]. This cycle is continuously executed 'k' times, allowing each unique fold a turn in the validation spotlight. My implementation of this approach makes sure the model's proficiency, measured through metrics such as accuracy, precision, recall, and F1 score, is not merely a consequence of fortunate data sampling. The following result was obtained after the K-fold Cross-Validation process. The process allowed to decrease overfitting reducing the precision from 1.0 to 0.9.

As this is not a major change in precision value, the model can be further improved to reduce chances of overfitting discussed in Chapter 5

(Precision: 0.99, Recall: 0.99, F1: 0.99, Loss: 0.004)

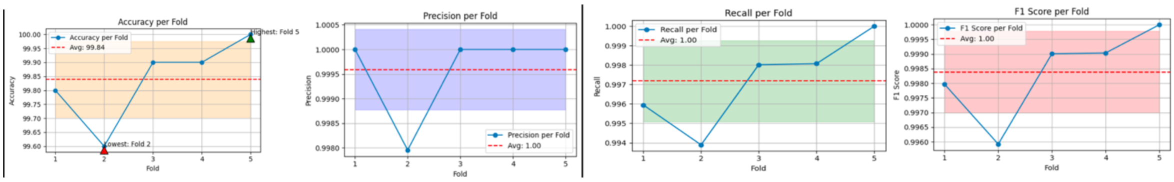


Fig 9 : Thumbnail for K-fold Cross Validation Performance Metrics with Standard Deviation. Please look at Appendix C.1 for a clearer/larger diagram

4.3.1.2 LIME (Local Interpretable Model-agnostic Explanations)

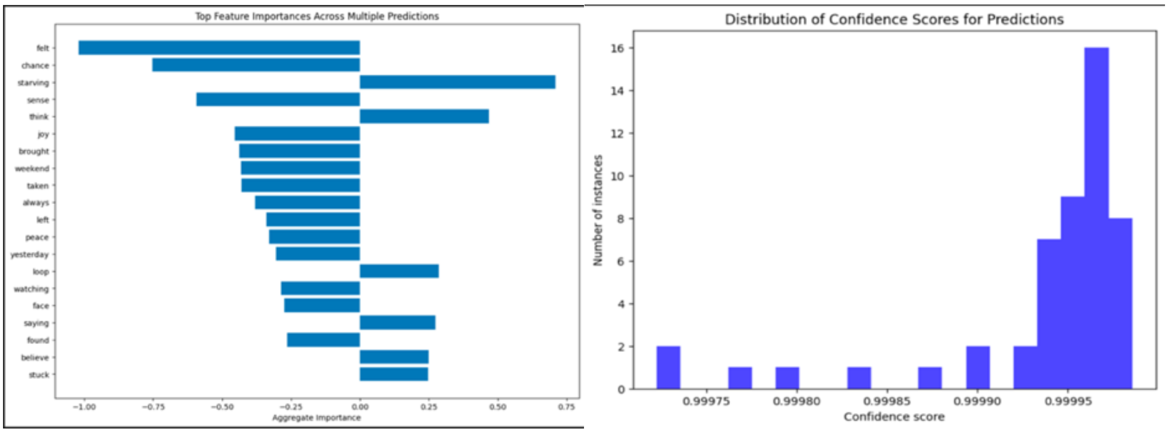


Fig 10 : Showing Top Feature Importance Across Multiple Predictions and Distribution of Confidence Scores for Predictions

To gain a deeper understanding of the model's decision-making process, I've employed LIME, which stands for Local Interpretable Model-agnostic Explanations. LIME is an algorithm that helps explain how a model that classifies or predicts outcomes works. It does this by simplifying the model's behaviour in an understandable way showing particular point of interest [75].

The first chart shows the importance of different words when the model is trying to decide if a text is 'Negative' or 'Positive'. Words that have bars stretching to the right are more important when the model thinks the text is 'Positive', and words that have bars going to the left are more important for 'Negative'. For example, words like "joy" and "peace" seem to be strong indicators for 'Positive', while words like "stuck" and "starving" are key for 'Negative'.

The second chart is about how sure the model is when it makes a guess. Most of the time, the model is really sure about its predictions, as we can see lots of instances with confidence scores close to 1 (which is the highest). There are a few times the model isn't as sure, shown by the smaller bars on the left of the chart, but those are less common. This tells us that the model generally predicts with high certainty.

4.3.1.3 Model Evaluation

After training, the model's performance must be evaluated on data it has never seen before, known as the test set. This evaluation is crucial as it is indicative of how well the model will perform when deployed in the real world [69]. Several metrics were used to assess the model's performance: Accuracy, Precision, Recall, F1-score, ROC AUC and ROC Curves.

Confusion Matrices: The model predicted 'Not Autistic' for 507 people, and they were actually not autistic. So, it got all of those right (no false positives). For 'Autistic', the model predicted that 493 people were autistic, and they were indeed autistic (no false negatives).

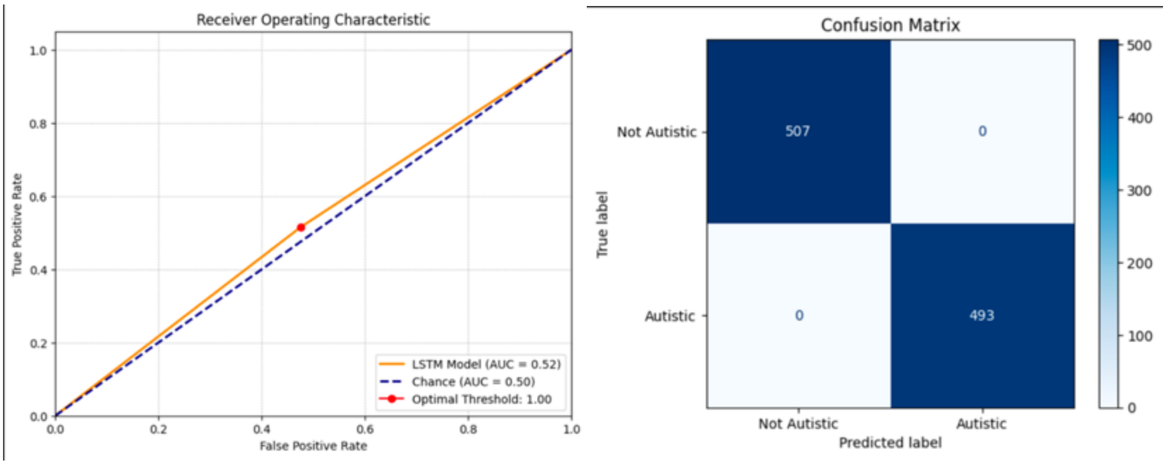


Fig 11 : ROC and Confusion Matrix

4.3.2 Eye Tracking

The model's performance varies across the two classes: Class 0 (the "negative" class) shows a precision of 0.77, meaning it correctly predicts class '0' 77% of the time, and an impressive recall of 0.92, capturing 92% of all actual class 0 instances. The F1-score of 0.84 indicates a strong balance between precision and recall, likely aided by the larger sample size of 171,402 instances in the dataset. Conversely, Class 1 (the "positive" class) demonstrates slightly higher precision at 0.78 but a significantly lower recall of 0.51, suggesting the model misses nearly half of the true class 1 instances. The F1-score of 0.62 for class 1 reflects a weaker performance in balancing precision and recall, which might be partly due to having fewer examples (98,830 instances) compared to class 0 (171,402 instances). The average accuracy across all folds during k-fold cross-validation was 0.76 showing that on average the model correctly predicts the outcome 76% of the time across different subsets of the data. This indicates that the model has a fairly good generalization when tested on different subsets of the data during the cross-validation process. The overall accuracy came to 0.77 or 77% correct predictions.

From the above we can conclude that the model is more effective at identifying the negative instances than the positive ones. The high recall for class 0 means the model is good at catching the negative instances but might be over-predicting them, potentially at the expense of class 1 accuracy. The model's ability to correctly identify positive instances (class 1) is significantly weaker, which is particularly concerning since class 1 reflects individuals with ASD. The data appears to be imbalanced, with class 0 having nearly twice the number of instances as class 1, which may contribute to the disparity in performance. To improve the model's predictive performance, especially for class 1 we should consider collecting more data for class 1, if possible or using resampling techniques to address class imbalance. If no more data is found, we could adjust class weights in the model training process. This has been explained in detail in Chapter 5, Future Work.

4.3.2.1 Evaluation Metrics

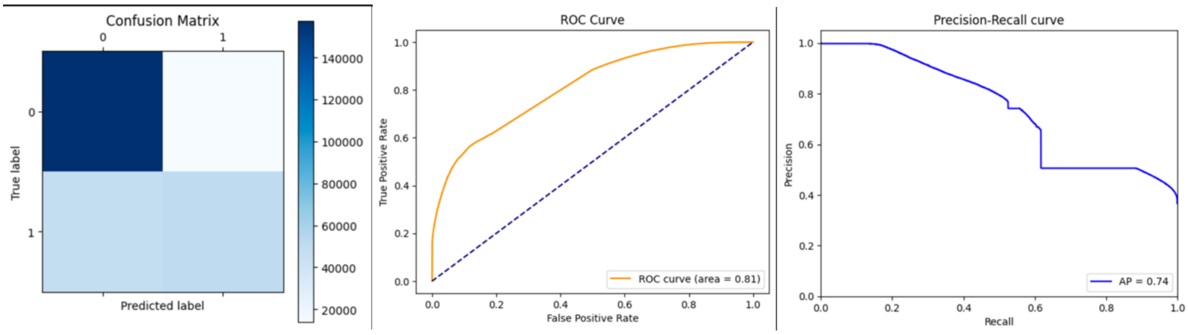


Fig 12: Confusion Matrix, ROC Curve, Precision-Recall curve

The confusion matrix shows a high number of true positives and true negatives, demonstrating the model's effectiveness in accurately identifying individuals as autistic or not, with minimal false positives and false negatives. The model's ability to accurately distinguish between the classes is further shown by the ROC curve, which has an area under the curve (AUC) of 0.81, indicating a strong capacity for separation. Additionally, the Precision-Recall curve highlights the model's high precision across various recall levels, supported by an Average Precision (AP) score of 0.74, suggesting that the model consistently delivers a high proportion of relevant results.

4.4 User Feedback shown on the Application

User feedback shown on the application was analysed against a benchmark dataset provided by Baron-Cohen et al., 2001 [12], which includes a diverse representation of individuals with conditions such as Asperger syndrome and high-functioning autism. This dataset also features balanced gender representation and includes a proportion of professionals from STEM fields, which allows for an accurate alignment and comparison of individual feedback. This process enhances the user experience by providing detailed explanations of personal questionnaire scores, offering users valuable insights into their interactions with the application.

Chapter 5
Discussion

5.1 Conclusion

This project has focused on the development of a healthcare application aimed at self-diagnosing Autism Spectrum Disorder (ASD) in adults, a group often overlooked in healthcare initiatives. By integrating advanced technologies like machine learning and artificial intelligence, the application promises to use diagnostic methods which provide deeper insights into ASD's behaviours. Its user-centric design includes features such as the eye-movement tracker and social interaction games, enhancing both the application's usability and its appeal to users and healthcare professionals alike.

Throughout the project, the collaborative and individual efforts have led to a robust platform that achieved our initial goals. This success was driven by the strategic use of agile methodologies and tools like GitHub, which enhanced flexibility and helped overcome technical challenges.

However, the project also uncovered areas needing further improvement. Although the application successfully merges various functionalities into a user-friendly interface, user testing revealed the need for more adjustments to fully meet the diverse needs of the ASD community. While the feedback on the interface was generally positive, it highlighted the necessity for making it more intuitive and accessible to all users.

5.1.1 Strengths

One of the major strengths of this project is its comprehensive approach to filling the gap in diagnostic tools for adults with Autism Spectrum Disorder (ASD). By using technologies like AI and machine learning, the project not only makes diagnoses more accurate but also deepens our understanding of ASD. The design of the application focuses on easy-to-use interfaces and engaging features that help keep users interested, which is very important for making sure the tool is used regularly and effectively. The teamwork throughout the project has also been crucial for developing a strong application. Each team member's unique skills and knowledge have added a lot to the project, helping to build a tool that is not only functional and secure but also pleasant to use. This team effort has sped up the development process and made sure the application addresses a wide range of user needs.

5.1.2 Limitations

Despite its successes, the project does have some weaknesses and limitations that might affect how well it works in the long run. A major issue is that the project depends on existing data sets, which might not fully capture the wide range of characteristics seen in adults with Autism Spectrum Disorder (ASD). This limitation could make it hard for the AI model to work well across the diverse group of adults with ASD. Additionally, the project's reliance on advanced technology also means that it needs a lot of ongoing support, updates, and maintenance, which could require a lot of resources. Moreover, the project is currently being used in a controlled setting, and its effectiveness in real-world situations hasn't been fully tested yet. Moving from a development environment to real-world use is often a big hurdle for many tech solutions, and this project is no exception.

5.2 Future Work – Shared Platform

5.2.1 AWS implementation

One significant hurdle was the lack of storage space provided by the AWS Free Tier. Since our application is still in the prototype phase, we opted to use AWS Free Tier to host it. Unfortunately, the storage space offered was insufficient, hindering our ability to continue hosting the application effectively. We encountered similar challenges with other hosting platforms we explored, such as Heroku and PythonAnywhere. As a result, we decided to host only the basic requirements of the application to create a prototype, which would represent a potential method for hosting the entire application in the future.

Additionally, we encountered problems with nginx when trying to connect our frontend, which was deployed using GitHub, to the backend deployed on the EC2 instance. Although CloudFront was set up for the instance, it only displayed the "Welcome to Nginx" homepage rather than our application. We considered connecting the backend to the frontend using an SSL certificate as an alternative method, but financial constraints prevented us from pursuing this option. Generating an SSL certificate was possible, but it would have been flagged as inauthentic on the server, so we decided not to continue with this process.

Furthermore, we have not implemented any firewall rules, which adds another layer of complexity and potential security issues to our project. Firewalls are crucial for regulating the traffic between the internet and the server hosting the application, providing a necessary barrier against unauthorized access and potential threats.

5.2.2 Backend Load testing

As mentioned earlier in Chapter 4, the presence of a maximum response time of 1434 ms suggests there are potential areas for optimization. Analysing slow transactions for optimization potential, possibly improving database queries or caching strategies. Further tests with higher user loads could also be beneficial to determine the system's scalability threshold. Other methods could include implementing real-time monitoring to capture and address performance dips immediately during unexpected traffic surges. By utilizing these insights, we can ensure that the backend platform remains robust and efficient, providing a reliable foundation for user interactions and future scalability.

5.3 Future Work – Autism Detector

1. General Practitioner feedback

The application was built using user feedback in the form of a google form and psychiatrist (based in India) feedback, as a demonstrative interview, taken into consideration but there are some factors to bear in mind: Due to time constraints, I was unable to get a UK General Practitioner feedback for the Autism Detector application. It is important to increase the authenticity of the project and follow health guidelines. To ensure this one of the most important tasks at hand would be to get feedback from a GP showing each feature of the application such that the resultant prototype is as close to a proper medical care technical solution.

2. Notes Tracking feature

During the interview with the psychiatrist, I learned that the note-tracking feature is generally used for patients with schizophrenia. However, further discussion revealed that it could also serve as a research tool for individuals with ASD. Since this feature tracks patterns such as repetition, language idiosyncrasies, detail orientation, and emotional expression, it could help users gain deeper insights into their behaviour, aiding their self-diagnosis process. Unfortunately, this interview happened late in the project, so I couldn't make adjustments based on these insights. I decided to maintain this feature as a research tool, as discussed in the interview.

3. Eye Tracking Training improvement

Due to the lower number of instances of Class 1 compared to Class 0, there is a potential bias towards the more frequent class. Balancing the class distribution through techniques such as oversampling Class 1 or under sampling Class 0 could enhance recall for Class 1. Modifying the weights of the classes in the loss function is recommended to increase focus on Class 1, thereby prioritizing the reduction of errors in the minority class. Additionally, implementing stratified k-fold cross-validation, as used in the other model for note tracking in the application, can strengthen the robustness and reliability of the model's performance evaluation.

4. Backend Load Testing – Individual Application

After completing load testing on a shared application backend, a valuable future step would be to apply similar testing methods to individual applications. This targeted approach can help identify specific performance bottlenecks and optimization opportunities within each application’s unique environment.

List of References

1. Yaneva, V., Eraslan, S., Yesilada, Y. and Mitkov, R., 2020. Detecting high-functioning autism in adults using eye tracking and machine learning. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(6), pp.1254-1261.
2. Muhle, R., Trentacoste, S.V. and Rapin, I., 2004. The genetics of autism. Pediatrics, 113(5), pp.e472-e486.
3. Murphy, C.M., Wilson, C.E., Robertson, D.M., Ecker, C., Daly, E.M., Hammond, N., Galanopoulos, A., Dud, I., Murphy, D.G. and McAlonan, G.M., 2016. Autism spectrum disorder in adults: diagnosis, management, and health services development. Neuropsychiatric disease and treatment, pp.1669-1686.
4. Pavlov, N., 2014. User interface for people with autism spectrum disorders. Journal of Software Engineering and Applications, 2014.
5. Department of Health (UK), “Basic Guidelines for People Who Commission Easy Read Information,” Department of Health, UK, 2009.
6. Khowaja, K. and Salim, S.S., 2013. A systematic review of strategies and computer-based intervention (CBI) for reading comprehension of children with autism. Research in Autism Spectrum Disorders, 7(9), pp.1111-1121.
7. Lin, C.S., Chang, S.H., Liou, W.Y. and Tsai, Y.S., 2013. The development of a multimedia online language assessment tool for young children with autism. Research in developmental disabilities, 34(10), pp.3553-3565.
8. Diagnostic and statistical manual of mental disorders: DSM-IV-TR American Psychiatric Publishing, Inc., Washington, DC (2000)
9. Kanner, L., 1943. Autistic disturbances of affective contact. Nervous child, 2(3), pp.217-250.
10. Whalon, K.J. and Hart, J.E., 2011. Children with autism spectrum disorder and literacy instruction: An exploratory study of elementary inclusive settings. Remedial and Special education, 32(3), pp.243-255.
11. Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H. and Baron-Cohen, S., 2015. Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. Molecular autism, 6, pp.1-12.
12. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. and Clubley, E., 2001. The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. Journal of autism and developmental disorders, 31, pp.5-17.

13. Huang, Y., Arnold, S.R., Foley, K.R. and Trollor, J.N., 2020. Diagnosis of autism in adulthood: A scoping review. Autism, 24(6), pp.1311-1327.
14. Lai, M.C. and Baron-Cohen, S., 2015. Identifying the lost generation of adults with autism spectrum conditions. The Lancet Psychiatry, 2(11), pp.1013-1027.
15. Lai, M.C., Lombardo, M.V., Ruigrok, A.N., Chakrabarti, B., Auyeung, B., Szatmari, P., Happé, F., Baron-Cohen, S. and MRC AIMS Consortium, 2017. Quantifying and exploring camouflaging in men and women with autism. Autism, 21(6), pp.690-702.
16. Jones, L., Goddard, L., Hill, E.L., Henry, L.A. and Crane, L., 2014. Experiences of receiving a diagnosis of autism spectrum disorder: A survey of adults in the United Kingdom. Journal of autism and developmental disorders, 44, pp.3033-3044.
17. Rutherford, M., McKenzie, K., Johnson, T., Catchpole, C., O'Hare, A., McClure, I., Forsyth, K., McCartney, D. and Murray, A., 2016. Gender ratio in a clinical population sample, age of diagnosis and duration of assessment in children and adults with autism spectrum disorder. Autism, 20(5), pp.628-634.
18. Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D. and Charman, T., 2006. Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). The lancet, 368(9531), pp.210-215.
19. Allison, C., Auyeung, B. and Baron-Cohen, S., 2012. Toward brief "red flags" for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. Journal of the American Academy of Child & Adolescent Psychiatry, 51(2), pp.202-212.
20. Embrace Autism Autism Spectrum Quotient. Available at: https://embrace-autism.com/autism-spectrum-quotient/ (Accessed: December 2023).
21. Booth, T., Murray, A.L., McKenzie, K., Kuenssberg, R., O'Donnell, M. and Burnett, H., 2013. Brief report: An evaluation of the AQ-10 as a brief screening instrument for ASD in adults. Journal of autism and developmental disorders, 43, pp.2997-3000.
22. Wilson, C.E., Roberts, G., Gillan, N., Ohlsen, C., Robertson, D. and Zinkstok, J., 2014. The NICE guideline on recognition, referral, diagnosis and management of adults on the autism spectrum. Advances in Mental Health and Intellectual Disabilities, 8(1), pp.3-14.
23. Hull, L., Mandy, W., Lai, M.C., Baron-Cohen, S., Allison, C., Smith, P. and Petrides, K.V., 2019. Development and validation of the camouflaging autistic traits questionnaire (CAT-Q). Journal of autism and developmental disorders, 49, pp.819-833.
24. Lai, M.C., Lombardo, M.V., Pasco, G., Ruigrok, A.N., Wheelwright, S.J., Sadek, S.A., Chakrabarti, B., MRC Aims Consortium and Baron-Cohen, S., 2011. A behavioral

comparison of male and female adults with high functioning autism spectrum conditions. PloS one, 6(6), p.e20835.

25. Bargiela, S., Steward, R. and Mandy, W., 2016. The experiences of late-diagnosed women with autism spectrum conditions: An investigation of the female autism phenotype. Journal of autism and developmental disorders, 46, pp.3281-3294.
26. Tierney, S., Burns, J. and Kilbey, E., 2016. Looking behind the mask: Social coping strategies of girls on the autistic spectrum. Research in Autism Spectrum Disorders, 23, pp.73-83.
27. Dean, M., Harwood, R. and Kasari, C., 2017. The art of camouflage: Gender differences in the social behaviors of girls and boys with autism spectrum disorder. Autism, 21(6), pp.678-689.
28. Cage, E., Di Monaco, J. and Newell, V., 2018. Experiences of autism acceptance and mental health in autistic adults. Journal of autism and developmental disorders, 48, pp.473-484.
29. Head, A.M., McGillivray, J.A. and Stokes, M.A., 2014. Gender differences in emotionality and sociability in children with autism spectrum disorders. Molecular autism, 5, pp.1-9.
30. Mandy, W., Chilvers, R., Chowdhury, U., Salter, G., Seigal, A. and Skuse, D., 2012. Sex differences in autism spectrum disorder: evidence from a large sample of children and adolescents. Journal of autism and developmental disorders, 42, pp.1304-1313.
31. Kopp, S. and Gillberg, C., 2011. The Autism Spectrum Screening Questionnaire (ASSQ)-Revised Extended Version (ASSQ-REV): An instrument for better capturing the autism phenotype in girls? A preliminary study involving 191 clinical cases and community controls. Research in developmental disabilities, 32(6), pp.2875-2888.
32. Ritvo, R.A., Ritvo, E.R., Guthrie, D., Ritvo, M.J., Hufnagel, D.H., McMahon, W., Tonge, B., Mataix-Cols, D., Jassi, A., Attwood, T. and Eloff, J., 2011. The Ritvo Autism Asperger Diagnostic Scale-Revised (RAADS-R): a scale to assist the diagnosis of autism spectrum disorder in adults: an international validation study. Journal of autism and developmental disorders, 41, pp.1076-1089.
33. Ritvo, R.A., Ritvo, E.R., Guthrie, D., Yuwiler, A., Ritvo, M.J. and Weisbender, L., 2008. A scale to assist the diagnosis of autism and Asperger’s disorder in adults (RAADS): A pilot study. Journal of Autism and Developmental Disorders, 38, pp.213-223.
34. Putnam, C. and Chong, L., 2008, October. Software and technologies designed for people with autism: what do users want?. In Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (pp. 3-10).

35. Tomczak, M.T., 2021. Employees with autism spectrum disorders in the digitized work environment: Perspectives for the future. Journal of disability policy studies, 31(4), pp.195-205.
36. Alabbas, N.A. and Miller, D.E., 2019. Challenges and assistive technology during typical routines: Perspectives of caregivers of children with autism spectrum disorders and other disabilities. International Journal of Disability, Development and Education, 66(3), pp.273-283.
37. Caldwell, T., 2020. A spectrum of tech: An integrated literature review of technologies to target social skills in students with autism spectrum disorders.
38. Fage, C., Consel, C.Y. and Balland, E., 2018. Tablet apps to support first school inclusion of children with autism spectrum disorders (ASD) in mainstream classrooms: A pilot study. Frontiers in psychology, 9, p.394873.
39. Wang, M. and Jeon, M., 2023. Assistive technology for adults on the autism spectrum: A systematic survey. International Journal of Human–Computer Interaction, pp.1-20.
40. Embrace Autism AQ-10. Available at: https://embrace-autism.com/aq-10/ (Accessed: December 2023).
41. Embrace Autism CAT-Q. Available at: https://embrace-autism.com/cat-q/ (Accessed: December 2023).
42. Embrace Autism RAADS-R. Available at: https://embrace-autism.com/raads-r/ (Accessed: December 2023).
43. Mozilla Developer Network, Cross-Origin Resource Sharing (CORS). Available at: https://developer.mozilla.org/en-US/docs/Web/HTTP/CORS (Accessed: December 2023).
44. Arisoy, E., Sainath, T.N., Kingsbury, B. and Ramabhadran, B., 2012, June. Deep neural network language models. In Proceedings of the NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT (pp. 20-28).
45. Hochreiter, S. and Schmidhuber, J. (1997) 'Long Short-Term Memory', Neural Computation, 9(8), pp. 1735-1780. Available at: https://direct.mit.edu/neco/article/9/8/1735/6109/Long-Short-Term-Memory (Accessed: December 2023).
46. Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, p.132306.
47. Schuster, M. and Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE transactions on Signal Processing, 45(11), pp.2673-2681.

48. Paperspace (no date) 'Advanced Recurrent Neural Networks: Bidirectional RNNs'. Available at: https://blog.paperspace.com/bidirectional-rnn-keras/#:~:text=A%20Bidirectional%20RNN%20is%20a,learning%20to%20past%20and%20present (Accessed: February 2024).
49. Cambria, E. and White, B., 2014. Jumping NLP curves: A review of natural language processing research. IEEE Computational intelligence magazine, 9(2), pp.48-57.
50. Ali Haider (June 23, 2021) 'Semantic Analysis Using Python'. Available at: https://medium.com/red-buffer/semantic-analysis-using-python-296639d082aa (Accessed: January 2024).
51. Suvrat Arora (9 Jan 2024) 'Sentiment Analysis Using Python'. Analytics Vidhya, 07 July. Available at: https://www.analyticsvidhya.com/blog/2022/07/sentiment-analysis-using-python/ (Accessed: February 2024).
52. Pipalia, K., Bhadja, R. and Shukla, M., 2020, December. Comparative analysis of different transformer based architectures used in sentiment analysis. In 2020 9th international conference system modeling and advancement in research trends (SMART) (pp. 411-415). IEEE.
53. Wittig, A. and Wittig, M., 2023. Amazon Web Services in Action: An in-depth guide to AWS. Simon and Schuster.
54. Mell, P. and Grance, T., 2011. The NIST definition of cloud computing.
55. Deo, R.C., 2015. Machine learning in medicine. Circulation, 132(20), pp.1920-1930.
56. Baştanlar, Y. and Özuysal, M., 2014. Introduction to machine learning. miRNomics: MicroRNA biology and computational analysis, pp.105-128.
57. Song, D.Y., Kim, S.Y., Bong, G., Kim, J.M. and Yoo, H.J., 2019. The use of artificial intelligence in screening and diagnosis of autism spectrum disorder: a literature review. Journal of the Korean Academy of Child and Adolescent Psychiatry, 30(4), p.145.
58. Beyond Autism (2022) 'Data Protection Policy'. Available at: https://www.beyondautism.org.uk/wp-content/uploads/2022/01/Data-Protection-Policy.pdf (Accessed: January 2024).
59. UK Government (2018) 'Data Protection Act 2018, Section 2'. Available at: https://www.legislation.gov.uk/ukpga/2018/12/section/2/enacted (Accessed: January 2023).
60. Richards, N. and Hudson, I., 2016. UK medicines regulation: responding to current challenges. British Journal of Clinical Pharmacology, 82(6), pp.1471-1476.
61. Caldwell, B., Cooper, M., Reid, L.G., Vanderheiden, G., Chisholm, W., Slatin, J. and White, J., 2008. Web content accessibility guidelines (WCAG) 2.0. WWW Consortium (W3C), 290(1-34), pp.5-12.

62. Haney, J.L., 2016. Autism, females, and the DSM-5: Gender bias in autism diagnosis. Social Work in Mental Health, 14(4), pp.396-407.
63. https://dbdiagram.io/d
64. Google form : https://docs.google.com/forms/d/e/1FAIpQLScnf7a70Da1gAfYZRSIJY6X_Z-nkqafZ0E8SoHDSBgiBBga3g/viewform?usp=sf_link
65. Previous year Architecture: https://drive.google.com/file/d/1K8AJfws8aYT-Og_pCWu16qPdOzBe1cUq/view
66. University of Leeds, Year, COMP2912 Software Engineering Principles, lecture notes, viewed 2022-23, <https://webprod3.leeds.ac.uk/catalogue/dynmodules.asp?Y=202021&M=COMP-2912>.
67. University of Leeds, Year, COMP1921 Programming Project, lecture notes, viewed 2022-23, <https://webprod3.leeds.ac.uk/catalogue/dynmodules.asp?Y=201819&M=COMP-1921>
68. University of Leeds, Year, COMP3611 Machine Learning, lecture notes, viewed 2023-24, <https://webprod3.leeds.ac.uk/catalogue/dynmodules.asp?Y=202223&M=COMP-3611>
69. University of Leeds, Year, COMP3631 Intelligent Systems and Robotics, lecture notes, viewed 2023-24, <https://webprod3.leeds.ac.uk/catalogue/dynmodules.asp?Y=202122&M=COMP-3631>
70. Previous year code Frontend: https://gitlab.com/comp3931-larks/comp3931-larks-ui-base
71. Previous year code Backend: https://gitlab.com/AJYC/comp3931-individual-project-larks-base
72. Eye Tracking data: https://www.kaggle.com/datasets/imtkaggleteam/eye-tracking-autism?select=Eye-tracking+Output
73. University of Leeds, COMP2011 Web Application Development, lecture notes, viewed 2022-23, <https://webprod3.leeds.ac.uk/catalogue/dynmodules.asp?Y=201819&M=COMP-2011>
74. Guyon, I., Saffari, A., Dror, G. and Cawley, G., 2010. Model selection: beyond the bayesian/frequentist divide. Journal of Machine Learning Research, 11(1).

75. Ribeiro, M.T., Singh, S. and Guestrin, C., 2016, August. " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).
76. Bone, D., Bishop, S.L., Black, M.P., Goodwin, M.S., Lord, C. and Narayanan, S.S., 2016. Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. Journal of Child Psychology and Psychiatry, 57(8), pp.927-937.
77. Bussu, G., Jones, E.J., Charman, T., Johnson, M.H. and Buitelaar, J.K., 2018. Prediction of autism at 3 years from behavioural and developmental measures in high-risk infants: a longitudinal cross-domain classifier analysis. Journal of Autism and Developmental Disorders, 48, pp.2418-2433.
78. Duda, M., Ma, R., Haber, N. and Wall, D.P., 2016. Use of machine learning for behavioral distinction of autism and ADHD. Translational psychiatry, 6(2), pp.e732-e732.
79. Duda, M., Haber, N., Daniels, J. and Wall, D.P., 2017. Crowdsourced validation of a machine-learning classification system for autism and ADHD. Translational psychiatry, 7(5), pp.e1133-e1133.
80. Duda, M., Haber, N., Daniels, J. and Wall, D.P., 2017. Crowdsourced validation of a machine-learning classification system for autism and ADHD. Translational psychiatry, 7(5), pp.e1133-e1133.
81. Levy, S., Duda, M., Haber, N. and Wall, D.P., 2017. Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. Molecular autism, 8, pp.1-17.
82. Thabtah, F., Kamalov, F. and Rajab, K., 2018. A new computational intelligence approach to detect autistic features for autism screening. International journal of medical informatics, 117, pp.112-124.
83. Wall, D.P., Dally, R., Luyster, R., Jung, J.Y. and DeLuca, T.F., 2012. Use of artificial intelligence to shorten the behavioral diagnosis of autism.
84. Tariq, Q., Daniels, J., Schwartz, J.N., Washington, P., Kalantarian, H. and Wall, D.P., 2018. Mobile detection of autism through machine learning on home video: A development and prospective validation study. PLoS medicine, 15(11), p.e1002705.
85. Liu, W., Li, M. and Yi, L., 2016. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. Autism Research, 9(8), pp.888-898.
86. Li, B., Sharma, A., Meng, J., Purushwalkam, S. and Gowen, E., 2017. Applying machine learning to identify autistic adults using imitation: An exploratory study. PloS one, 12(8), p.e0182652.

87. Anzulewicz, A., Sobota, K. and Delafield-Butt, J.T., 2016. Toward the Autism Motor Signature: Gesture patterns during smart tablet gameplay identify children with autism. Scientific reports, 6(1), p.31107.
88. Crippa, A., Salvatore, C., Perego, P., Forti, S., Nobile, M., Molteni, M. and Castiglioni, I., 2015. Use of machine learning to identify children with autism and their motor abnormalities. Journal of autism and developmental disorders, 45, pp.2146-2156.
89. Markham, S., 2018. Diagnosis and treatment of ASD in women in secure and forensic hospitals. Advances in Autism, 5(1), pp.64-76.
90. Russell, G., Stapley, S., Newlove-Delgado, T., Salmon, A., White, R., Warren, F., Pearson, A. and Ford, T., 2022. Time trends in autism diagnosis over 20 years: a UK population-based cohort study. Journal of Child Psychology and Psychiatry, 63(6), pp.674-682.
91. Burton, A., McKinlay, A., Aughterson, H. and Fancourt, D., 2023. Impact of the COVID-19 pandemic on the mental health and well-being of adults with mental health conditions in the UK: a qualitative interview study. Journal of Mental Health, 32(6), pp.1040-1047.
92. Gantt Chart : https://docs.google.com/spreadsheets/d/1ZdVB8XFq_UupuYN9wDT9vikDBnrV8pJfa4vTyHdRHRo/edit#gid=0
93. Individual Gantt Chart : https://docs.google.com/spreadsheets/d/1EZkVD57nPOLZ5vIHCWtxlcQXrQJt5PZQc0Mq9Xd9eKs/edit#gid=0
94. Amazon Web Services 'Content Delivery Network (CDN)'. Available at: https://aws.amazon.com/caching/cdn/#:~:text=A%20Content%20Delivery%20Network%20(CDN,distributed%20set%20of%20caching%20servers (Accessed: February 2024).
95. Turing 'Guide on Word Embeddings in NLP'. Available at: https://www.turing.com/kb/guide-on-word-embeddings-in-nlp (Accessed: January 2024).
96. Glove : https://nlp.stanford.edu/projects/glove/
97. TutorialsPoint 'Keras Dense Layer'. Available at: https://www.tutorialspoint.com/keras/keras_dense_layer.htm (Accessed: March 2024)
98. Analytics Vidhya (2021) 'Binary Cross-Entropy / Log Loss for Binary Classification'. Available at: https://www.analyticsvidhya.com/blog/2021/03/binary-cross-entropy-log-loss-for-binary-classification/ (Accessed: March 2024)

Appendix A
Self-appraisal

A.1 Critical self-evaluation

In this project, I aimed to develop an AI-based web application that supports self-diagnosis and tracking of Autism Spectrum Disorder (ASD) in adults. Initially, the project began with research into autism and its impacts. At first, I struggled to find existing tools or solutions that could serve as references for creating specific features. My lack of experience in this field made it difficult to locate suitable datasets for model training. After deciding on the features, I wanted to implement, I faced a significant challenge as there were no readily available datasets for the note-tracking feature. So, I had to create synthetic data based on characteristics typically found in individuals with ASD, which doubled my workload and disrupted my planned schedule.

As for the group dynamics, there were no significant difficulties except during the deployment process. None of us had prior experience with AWS or other deployment platforms like Heroku. I wanted to ensure the security of my individual application, given that it handles personal information, so I chose to host the backend on AWS. This decision was challenging due to time constraints and limited financial resources, which ultimately prevented me from fully hosting the application. Additionally, I encountered issues integrating the backend with the frontend, which was hosted on GitHub by one of the group members (Archie). My lack of experience and the limited time frame affected my ability to resolve this integration fully.

Another setback was the interview done late; I should have conducted the interview earlier in the project to allow time for necessary adjustments to the features. I also should have anticipated potential risks and limitations associated with each project feature to ensure I could achieve all my goals. This experience has taught me the importance of timely stakeholder engagement and risk management in project planning.

A.2 Personal reflection and lessons learned

One of the most significant lessons I learned is the importance of engaging with stakeholders early in the project. The late-stage feedback from the psychiatrist was highly beneficial, yet it came too late to influence the project's direction significantly. In future projects, I plan to involve industry experts and potential users from the beginning to ensure that the features developed are relevant and based on real-world needs.

The initial phase of the project revealed gaps in my knowledge about autism and available datasets, which hindered my progress. This experience underscored the necessity of thorough research and preparation before diving into the development phase. Moving forward, I will allocate more time to understand the domain fully, identify resources, and evaluate existing technologies related to any project. Coming into this project, I had limited knowledge of data training techniques, which are crucial for AI development. Coming from a background in Saudi Arabia, where exposure to psychological advancements and issues is not as prevalent, this project was particularly enlightening. It allowed me to delve into the realm of psychological health technologies, broadening my understanding of how technology can intersect with mental health to provide innovative solutions.

This project was an opportunity to learn about various methodologies, including K-fold Cross Validation, LIME, etc, which helped me understand how to train models more effectively and evaluate their performance rigorously. This was a significant step in my learning curve, enhancing my ability to handle complex datasets and training scenarios. The technical challenges, especially deploying the application on AWS and integrating backend and frontend services, pushed me to learn new technologies under pressure. This has taught me the value of resilience and adaptability in learning, traits that will undoubtedly aid me in future projects.

Apart from data training, I was introduced to various other technical aspects I was previously unfamiliar with, including eye movement tracking technologies and advanced NLP techniques. Learning these new technologies was challenging but ultimately rewarding, providing me with a broader toolkit for future endeavours. In conclusion, this project did not merely result in the development of a technical solution; it was a profound learning experience that expanded my understanding of an important healthcare domain. The insights gained from both the technical development and project management aspects of the project are invaluable for my career development, ensuring that my future endeavours are both impactful and well-informed.

A.3 Legal, social, ethical and professional issues

A.3.1 Legal issues

When developing a health-related AI application, which assists in self-diagnosis and monitoring of Autism Spectrum Disorder (ASD), there are several legal concerns I needed to consider. These concerns ensure that the application meets legal standards and respects the privacy and rights of its users. One of the main legal issues revolves around data privacy and protection. In the UK, this means complying with the General Data Protection Regulation (GDPR) [58] and the Data Protection Act 2018 [59]. These laws require that we handle user data very carefully. We must make sure that users know what data we collect and why, and they must agree to it explicitly. This makes the security of the data crucial as well. We have to use strong encryption when storing and sending data to protect it from unauthorized access.

Another big issue is about medical compliance and liability. If the app is seen as a medical device, we need to follow strict regulations set by the UK Medicines and Healthcare products Regulatory Agency (MHRA). The role of the MHRA is to regulate medicines, medical devices and blood components for transfusion in the UK [60]. This ensures that the app is safe and does what it claims to do without harming users. It's also important to remind users that the app isn't a replacement for professional medical advice. To do this I have implemented clear disclaimers that the app is for informational purposes and should not be used as the sole basis for diagnosing ASD. In conclusion, while the development of the application involves complex technological tasks, understanding and implementing the legal and ethical requirements is equally crucial.

A.3.2 Social issues

Addressing the social issues including accessibility and inclusivity emerged as primary concerns, ensuring the application is user-friendly for everyone, regardless of their technological proficiency or physical abilities. This involves making the interface intuitive and ensuring that all functionalities are accessible to individuals with disabilities, trying to follow guidelines like the Web Content Accessibility Guidelines (WCAG) [61].

Another crucial aspect is the potential for dependency on the application. Users might rely heavily on the app for insights into ASD, possibly at the expense of seeking professional medical advice. To mitigate this, I've integrated frequent reminders within the app encouraging users to consult healthcare professionals for comprehensive assessments. This is crucial for ensuring that the app complements professional advice rather than replacing it.

Privacy issues are particularly sensitive given the personal nature of the data handled. Ensuring robust data protection measures and maintaining transparency about data usage are foundational to building trust. I've prioritized clear communication about how data is secured and the purposes for which it is used, helping users feel more secure about their privacy.

Stigma associated with ASD is another significant concern. It was imperative to approach this aspect with sensitivity to avoid reinforcing stereotypes. The content within the app was carefully curated to portray ASD accurately and respectfully, aiming to educate and foster understanding rather than perpetuate stigma. The risk of bias in AI algorithms was a pressing issue, given the potential for these biases to influence the app’s assessments. Ensuring that the algorithms were trained on diverse datasets and continuously testing for bias became a routine part of the development process.

In addressing the social concerns of the AI application, gathering diverse data to train the AI accurately presented unique challenges, particularly the concept of "masking," where individuals with ASD may conceal their symptoms in social settings. This behaviour is often more pronounced in women, who are historically underdiagnosed. Emphasizing this aspect aims to provide insights that might help reduce misdiagnosis and support a demographic historically overlooked in ASD research [62]. However, this focus also raises important ethical considerations about the representation and fairness of the AI training process. To mitigate the risks of bias, I have taken steps to include as broad a spectrum of participants as possible in the training data, with an emphasis on including more data from women who exhibit masking behaviours. This has been explained in detail in the next section, A.3.3 Ethical Issues.

A.3.3 Ethical issues

Following the exploration of legal and social issues, the ethical concerns in the development of my AI application also demand careful consideration. The potential ethical implications are profound, given the sensitive nature of the health data involved and the application's role in influencing user perceptions and behaviors regarding Autism Spectrum Disorder (ASD).

Firstly, fairness and bias in AI algorithms are big concerns. It's essential that the AI does not continue existing biases or introduce new ones, which could negatively impact certain user groups. This has been a significant focus during the development process, particularly because of the application's emphasis on the concept of "masking," a behaviour observed in individuals with ASD. Masking is often more pronounced in women, leading to underdiagnosis. Ensuring that the AI algorithms account for these differences responsibly requires a nuanced approach to training data curation and algorithm testing.

Transparency is another critical ethical issue. Users must understand what data the app collects, how it is used, and the limitations of the AI's capabilities. I have prioritized transparency by providing clear, accessible information about the AI processes and data usage within the app. This not only builds trust but also empowers users to make informed decisions about their engagement with the app. Privacy and data security are intertwined with ethical considerations. Given the application's handling of sensitive health information, robust measures are in place to protect user data against unauthorized access and breaches. Adherence to GDPR and other privacy regulations underpins every aspect of data handling within the app, ensuring that users' rights to privacy are respected and protected.

A.3.4 Professional issues

In developing my AI application for monitoring Autism Spectrum Disorder (ASD), adhering to professional standards and practices has been very important. This includes maintaining a high level of professionalism in every interaction with stakeholders, including users, healthcare professionals, and regulatory bodies. Ensuring that all development practices meet industry standards for software development and health technology at the prototype stage is crucial. This includes rigorous testing and documentation to deliver a reliable and effective prototype. Upholding these professional standards ensures the application not only meets technical and ethical expectations but also aligns with the professional norms of the healthcare and technology sectors, fostering trust and credibility among all stakeholders involved.

Appendix B

External Materials

B.1 Previous year Architecture Diagram

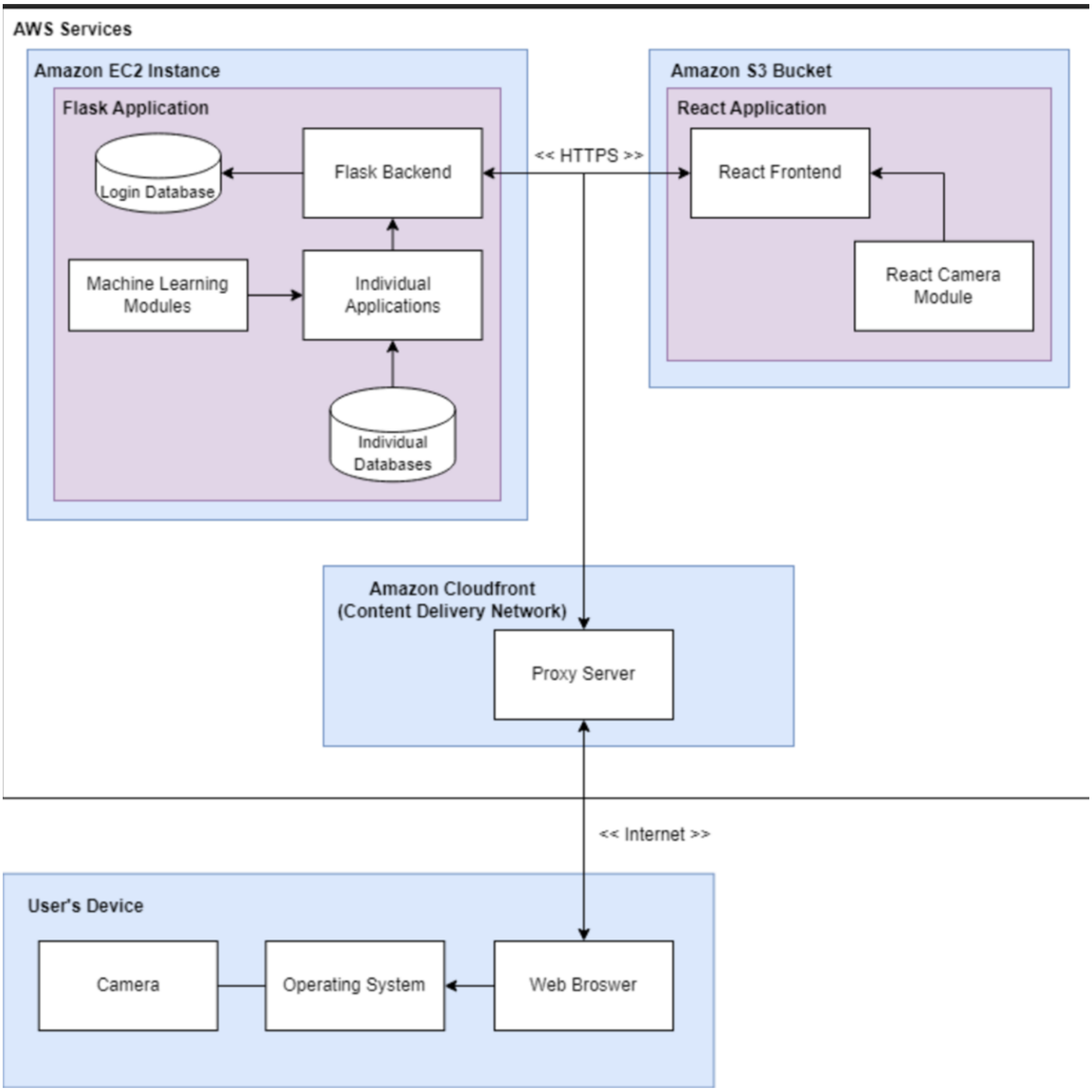


FIG : Previous year Architecture diagram

B.2 Summary of studies using AI technology with existing ASD assessments [57]

Author	Sample size	Mean age	Date type	Method	AUC (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Bone et	1264	6.7-	ADI-R	SVM	-	89.2 86.7	59 53.4	-

Author	Sample size	Mean age	Date type	Method	AUC (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
al. [76]	(ASD) 462 (non-ASD)	15.9 yrs 6.6-17.2 yrs	SRS					
Bussu et al. [77]	32(ASD) 43 (atypical) 86 (typical)	8.1 m (visiti) 14.5 m (visit2) 25.4 m (visit3) 38.4 m (visit4)	MSEL VABS AOSI	SVM	69.2 (8 m) 70.8 (14 m)	68.8 (8 m) 60.7 (14 m)	64.4 (8 m) 67.5 (14 m)	66.4 (8 m) 64.4 (14 m)
Duda et al. [78]	2775 (ASD) 150 (ADHD)	8.1 yrs (ASD) 11.3 yrs (ADHD)	SRS	SVC, LDA, CL, LR, RF, DT	93.3 - 96.5	-	-	-
Duda et al. [79]	248(ASD) 174 (ADHD)	8.2 yrs (ASD) 10.4 yrs (ADHD)	SRS	SVC, CL, LR, LDA	82-89	-	-	-
Kosmicki et al. [80]	1451 ASD (M2) 348 non-ASD (M2)	68 m ASD (M 2) 37 m non-ASD	ADOS (M2, M3)	SVM, ADTree, FT, LR, NBT,		96.7-99.7 (M2) 96.1-100 (M3)	96.5-98.6 (M2) 87.1-98.9 (M3)	

Author	Sample size	Mean age	Date type	Method	AUC (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
	2434 ASD (M 3) 307 non-ASD (M 3)	(M2) 111 m ASD (M3) 108 m non-ASD (M 3)		RF				
Levy et al. [81]	1319 ASD (M 2) 70 non-ASD (M 2) 2870 ASD (M 3) 273 non-ASD (M 3)	83 m ASD (M 2) 60 m non-ASD (M 2) 116 m ASD (M 3) 109 m non-ASD (M 3)	ADOS (M2, M3)	LR, LDA, SVM	93 (M2) 95 (M3)	98 (M2) 95 (M3)	58 (M2) 87 (M3)	78 (M2) 90 (M3)
Thabtah et al. [82]	707(ASD) 393 (non-ASD)	6.3 yrs 14.1 yrs 29.7 yrs	AQ	CI		80-87.3 80.95-82.54	80 80 90	
Wall et al. [83]	2867 (ASD) 92 (non-ASD)	8.06-8.75 yrs (ASD)	ADI-R	ADTree	-	-	93.8-99	99.9-100

Author	Sample size	Mean age	Date type	Method	AUC (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
		9.24-9.75 yrs (non-ASD)						

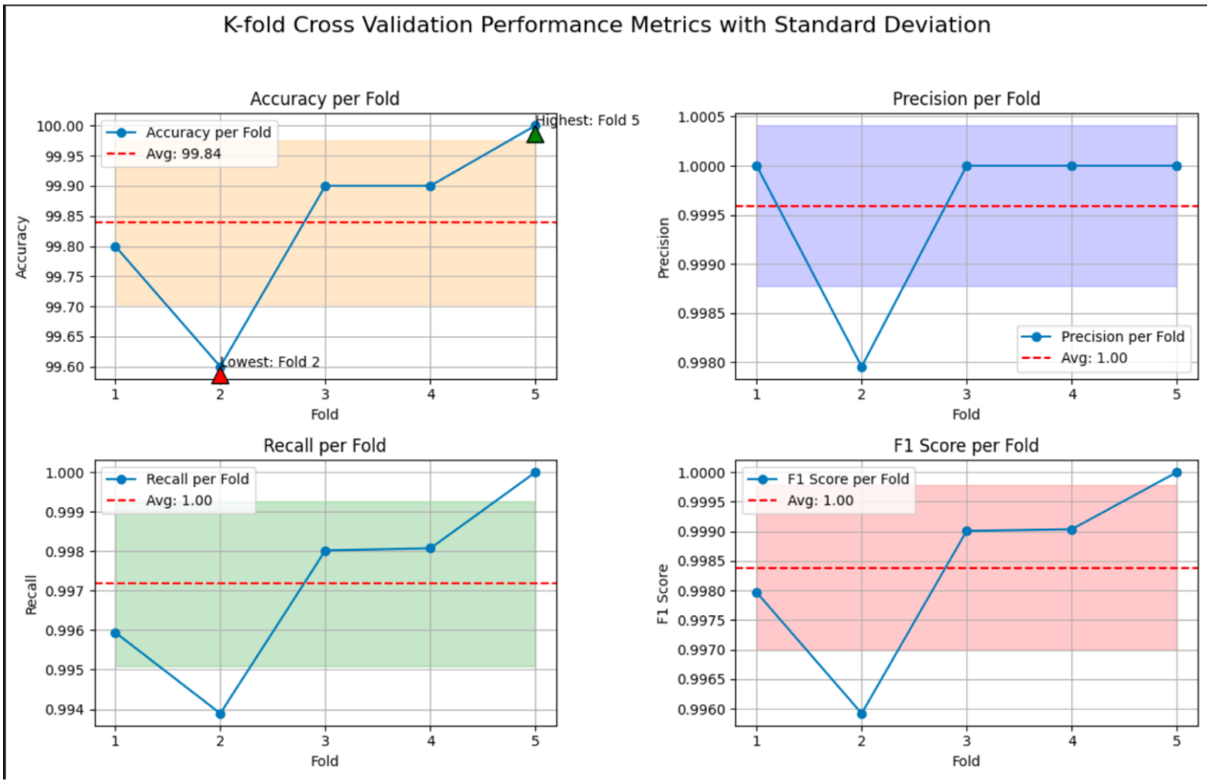
Summary of studies using AI technology with existing ASD assessments [57]

Author	Sample size	Mean age	Date type	Method	AUC (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Tariq et al. [84]	116 (ASD) 46 (TD)	4.10yrs (ASD) 2.11yrs (TD)	Behavioral features	ADTree, SVM, LR, RK, LİSVM Sparse 5 feature LR classifier	89-92	90-100	1.13-100	94-100
Liu et al. [85]	29(ASD) 29 (TD)	7.9 yrs (ASD) 7.86 yrs (TD)	Eye-tracking	SVM	89.63	93.1	86.21	88.51
Li et al. [86]	14(ASD) 16 (TD)	32 yrs (ASD) 29.31 yrs (TD)	Hand movement	NB, SVM, RF, DT	-	57.1-85.7	68.8-87.5	66.7-86.7

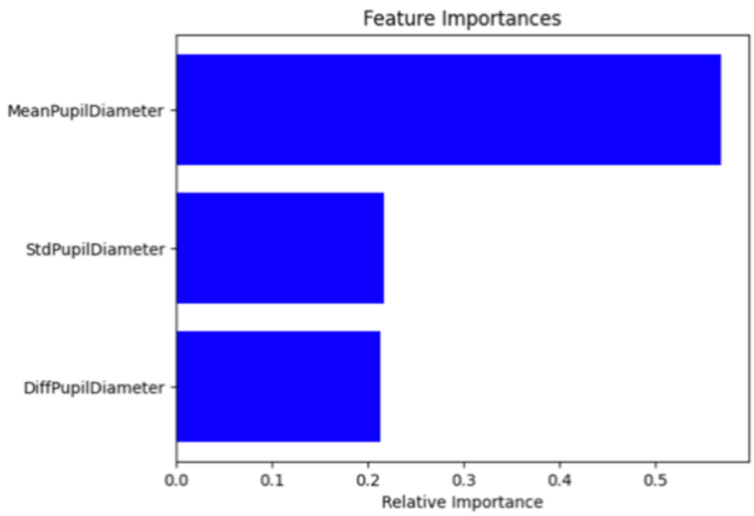
Author	Sample size	Mean age	Date type	Method	AUC (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Anzulewicz et al. [87]	37(ASD) 45 (TD)	4.5 yrs (ASD) 4.7 yrs (TD)	Hand movement	RF, RGF	88.1 - 93.2	76-83	67-88	-
Crippa et al. [88]	15(ASD) 15 (TD)	3.5yrs (ASD) 2.6yrs (TD)	Upper-limb movement	SVM	-	82.2-100	89.1-93.8	84.9-96.7

Appendix C
Additional Content

C.1 K-Fold Cross Validation Performance Metrics with Standard Deviation



C.2 EVALUATION METRIC – EYE TRACKING



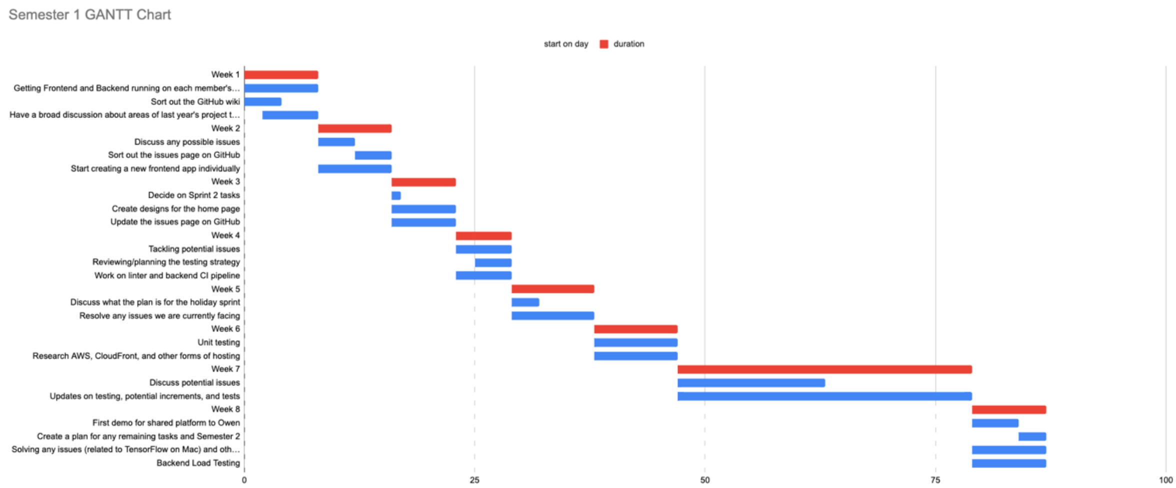
The Feature Importance chart reveals the relative significance of each feature used in the prediction. The mean pupil diameter is shown to be the most influential feature, followed by the standard deviation of pupil diameter and the absolute difference in pupil diameter. These insights into feature importance are critical for understanding the factors that most impact the model's predictions.

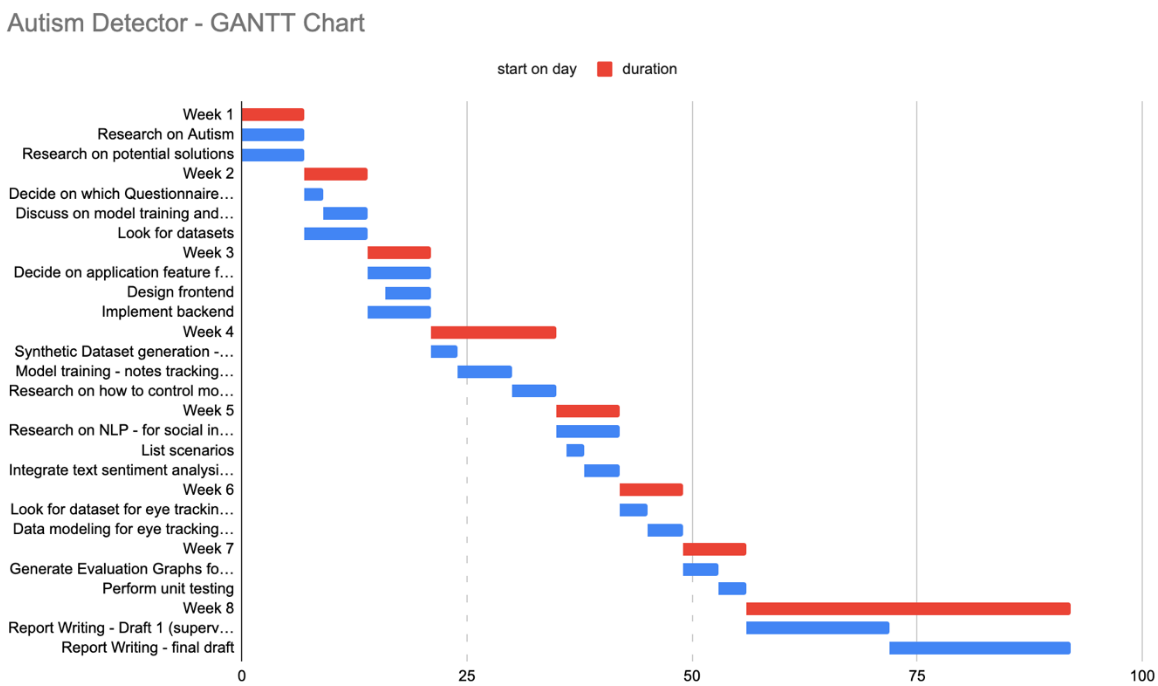
C.3 Risk Mitigation Table

Risk	**Impact**	**Mitigating Strategy**
Lack of experience with certain technology (ReactJS, AWS, ESlint, Jest, etc.)	Delays in development, potential for suboptimal implementations	Conduct pre-project training sessions; utilize online resources and tutorials; assign mentorship roles within the team where experienced members can guide others.
Ensuring previous year's code works on all machines (Mac and Windows)	Potential for significant delays if compatibility issues arise	Perform initial cross-platform compatibility tests; using machines used by each group member.
No previous experience working in separated frontend and backend	Challenges in integration, inefficiencies in development	Organize knowledge sharing sessions on best practices for API design and frontend-backend integration; encourage small, cross-functional teams to enhance understanding across layers.
Unforeseen technology requirements. Abrupt changes cause slowdown, or worst-case pivoting of the project. Errors in code, breaking application	Project delays, increased workload, potential for project failure	Implement an agile development approach with regular check-ins to adapt quickly to new requirements; maintain a robust version control system to manage changes effectively.
Unforeseen functional limitations	Inability to meet project objectives,	Conduct thorough architectural review and planning sessions with all team members;

Risk	**Impact**	**Mitigating Strategy**
caused by bad deployment architecture	need for significant redesign	engage in peer review of architectural decisions.
Pip install requirements not managed properly. Separate installations don't function well together and cause errors.	Project dependencies cause errors, leading to delays and potential for failure	Use virtual environments (e.g., Python's venv) to manage dependencies cleanly; maintain a well-documented requirements.txt file; regularly update and test dependency installations.
macOS having different instalment requirements.	Compatibility issues, leading to unexpected behaviour or failures on certain versions	Testing done on 2 MacOS machines owners by the group members.
Difference in opinion of group members	Delays, potential for conflict and reduced morale	Establish clear communication channels and conflict resolution strategies; implement a decision-making process that includes consensus-building or democratic voting.
Group members' personal issues (such as getting sick, emergencies, etc.)	Reduced manpower, potential for missed deadlines	Plan for contingencies by building buffer time into the project schedule; encourage cross-training among team members to cover essential tasks in absence.
No experience hosting on AWS	Difficulties in deployment, potential downtime	Leverage AWS training resources and documentation; start with simple deployments to gain familiarity before moving to more complex configurations.

C.4 Google Spreadsheet for Project Workflow [92, 93]





C.5 Database Schema – Individual Application Autism Detector

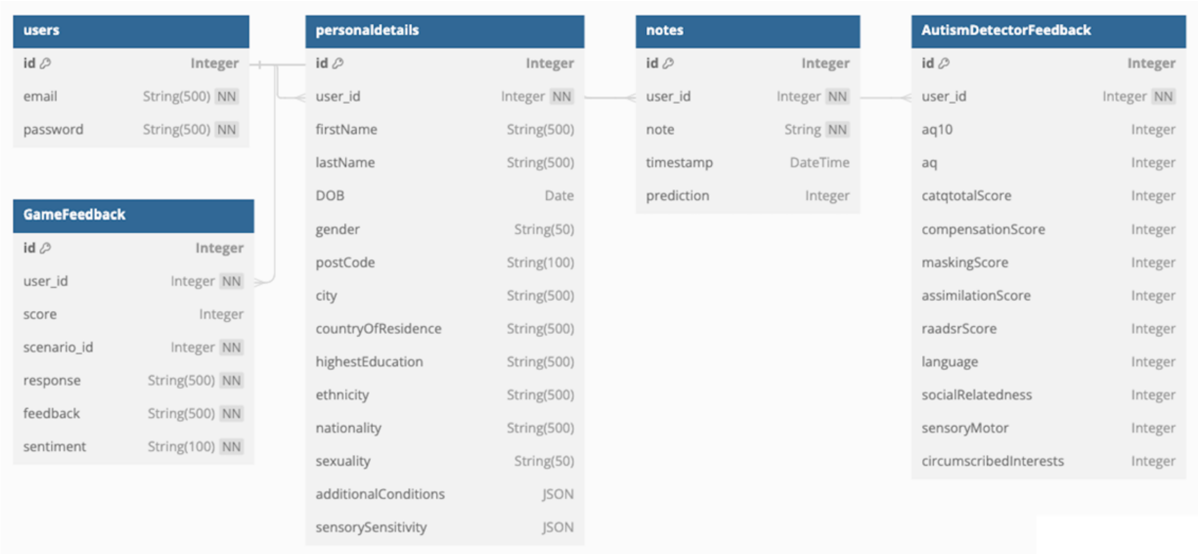
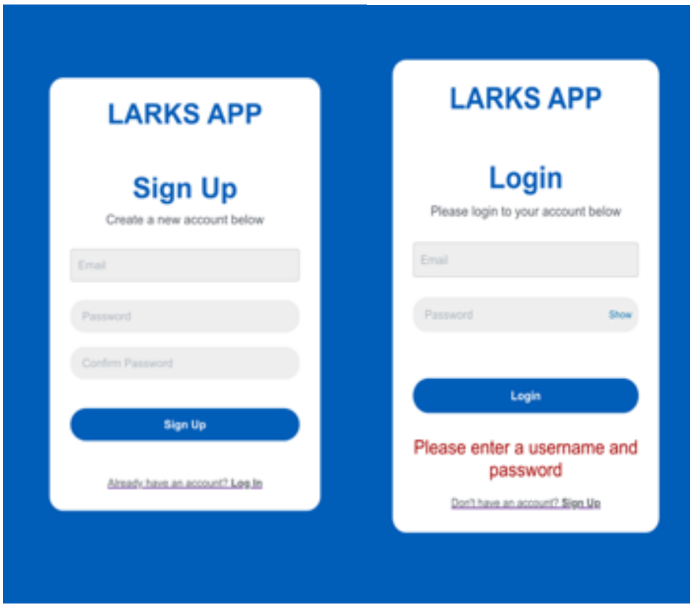


Fig: Database Schema – created using dbdiagram [63]

C.6 Login and Home page




LARKS APP
Apps
Logout
Skin Scan
dipstik
PARALYSIS ANALYSIS
Autism Detector
DepressiLess


Personal Details
Eye Tracking
Social Interaction Game
Test
Feedback
Tracking & Notes
Welcome 7755@gmail.co!
Hi! Would you like to talk to me?
Yes
Always
Unique
Totally
Intelligent
Sometimes
Mysterious
Take a Questionnaire
Try the Eye Movement Tracking
Click to Chat

Appendix D
Contribution Table

The five members of the team are as follows:

1. Archie Adams [ed199aa]
2. Ayesha Rahman [sc21ar]
3. Chien-Wei Tung [fy19cwt]
4. Marilena Manoli [sc21m2m]
5. Natalie Leung [sc21nl]

Group Member	Task	Area
All	Review previous year platform, code and wiki documentation	Frontend, Backend and GitHub
Archie	Merged the old two repositories with non-linear history while retaining all previous commits and git history.	Repository
Archie	Set up the wiki and copied across the legacy files.	Repository
Archie	Set up branch protection and pull request rules	Repository
Archie	Cleaning up the repository by removing build artefacts, setting proper config and tidying code.	Backend and Frontend
Archie	Refactor old frontend core into index, Routes, and App components in line with industry standards.	Frontend
Archie	Moved legacy apps into their own section and folders in the frontend.	Frontend
Archie	Refactor App.css to be multiple component level .scss files.	Frontend

Group Member	Task	Area
Archie	Replace instances of legacy developer's names in the code with their projects name	Frontend and Backend
Archie	Removed databases from being committed in the repo	Backend
Archie	Refactor most shared components to use typescript	Frontend
Archie	Fixed JWT authentication on signin	Backend and Frontend
Archie	Add error page component	Frontend
Archie	Implement frontend testing	Frontend
Archie	Unit test for error page component	Frontend
Archie	Unit test for header component	Frontend
Archie	Added eslint to frontend	Frontend
Archie	Implemented frontend CI to run unit tests, linter and app build on PR	Repository
Archie	Refactor views.py into endpoints.py, removing bugs and insecure code from login and signup endpoints	Backend
Archie	Implemented frontend CD to automatically deploy the frontend on github pages	Frontend
Archie	Added frontend unit test coverage command	Frontend
Archie	Implemented backend CI to run unit tests on PR	Backend
Archie	Rearchitected backend to use an application factory so duplicated testing tables in the database did not have to	Backend

Group Member	Task	Area
	be used.	
Archie	Rewrote backend unit tests to be more granular and concise and use a new in memory instance of the app and database per unit test.	Backend
Ayesha	Frontend Shared home re-design and implementation	Frontend
Ayesha	Setting up an EC2 instance - AWS	Backend
Ayesha	Connecting and installing Flask backend requirements for AWS	Backend
Ayesha	Setting up CloudFront - AWS	Backend to Frontend Integration
Ayesha	nginx Configuration - AWS	Backend to Frontend Integration
Ayesha	Changes to frontend .env file for deployed backend testing	Frontend
Ayesha	Testing deployed backend on different devices (Mac, Windows, Linux, iPhone, Android)	Backend
Ayesha	GitHub wiki documentation for AWS	GitHub wiki
Ayesha	Unit tests for SignIn page component	Frontend
Ayesha	Project Workflow Gantt Chart - Semester 1 Group work	GitHub wiki
Ayesha	Risk Mitigation Table documentation	GitHub wiki

Group Member	Task	Area
Ayesha	Data Security documentation	GitHub wiki
Ayesha	Backend load testing - using Artillery	Backend
Ayesha	Load testing documentation	GitHub wiki
Natalie	Route test for the shared platform, EaseMind and previous year's app.	Frontend
Natalie	"show" button for login page	Frontend
Natalie	Fixed button sizes on header	Frontend
Natalie	Prevent creation of same email address.	Backend
Natalie	Merging previous year's meeting notes.	Repository
Natalie	Validation for password on sign up page.	Frontend
Marilena	SignUp test for shared platform	Frontend
Marilena	Additions and modifications to requirements-m1working.txt file to ensure the code is compatible for M1 chip	Repository
Marilena	Restructuring of the wiki, adding new sections for minutes, aims and objectives etc.	GitHub Wiki
Marilena	Keeping track of minutes for all meetings held throughout the development of the project	GitHub Wiki
Marilena	Research on how to deploy app to AWS/ Heroku	Backend, Frontend

Group Member	Task	Area
Chien-Wei	Camera Jest tests for shared platform	Frontend
Chien-Wei	Categorize wiki documents by adding a sidebar	GitHub Wiki
Chien-Wei	Create wiki document for simulating React on mobile phones	GitHub Wiki
Chien-Wei	Update README for Jest usage	Repository

Appendix E

Backend Load Testing Results

Table 1: Time, HTTP Codes, and Request Rates

Time Period	HTTP Codes 200	Downloaded Bytes	HTTP Request Rate	HTTP Requests
20:27:30	21	4598	3/sec	22
20:27:40	36	7883	7/sec	37
20:27:50	60	13135	6/sec	60
20:28:00	60	13140	6/sec	60
20:28:10	66	14450	10/sec	67
20:28:20	86	19015	9/sec	88
20:28:30	88	19360	9/sec	89
20:28:40	94	20673	12/sec	95
20:28:50	120	26274	12/sec	120
20:29:00	120	26277	12/sec	120
20:29:10	124	27242	15/sec	125
20:29:20	150	32847	15/sec	150
20:29:30	150	32845	15/sec	150
20:29:40	150	32841	15/sec	150
20:29:50	150	32847	15/sec	150
20:30:00	150	32845	15/sec	150
20:30:10	150	32844	15/sec	150
20:30:20	150	32848	15/sec	150
20:30:30	150	32847	15/sec	150
20:30:40	150	32847	15/sec	150
20:30:50	150	32846	15/sec	150
20:31:00	150	32845	15/sec	150

20:31:10	150	32849	15/sec	150
20:31:20	150	32846	15/sec	150
20:31:30	150	32845	15/sec	150
20:31:40	150	32843	15/sec	150
20:31:50	150	32844	15/sec	150
20:32:00	150	32849	15/sec	150
20:32:10	150	32842	15/sec	150
20:32:20	150	32842	15/sec	150
20:32:30	150	32843	15/sec	150
20:32:40	150	32845	15/sec	150
20:32:50	150	32842	15/sec	150
20:33:00	150	32841	15/sec	150
20:33:10	150	32846	15/sec	150
20:33:20	150	32847	15/sec	150
20:33:30	150	32838	15/sec	150
20:33:40	150	32844	15/sec	150
20:33:50	150	32842	15/sec	150
20:34:00	150	32847	15/sec	150
20:34:10	150	32844	15/sec	150
20:34:20	150	32846	15/sec	150
20:34:30	55	11581	15/sec	47
Summary	5580	1221815	8/sec	5580

Table 2: Response Times

Time Period	**Min Response Time (ms)**	**Max Response Time (ms)**	**Mean Response Time (ms)**	**Median Response Time (ms)**	**P95 Response Time (ms)**	**P99 Response Time (ms)**

20:27:30	3	369	220.1	320.6	347.3	347.3
20:27:40	2	343	210	308	327.1	327.1
20:27:50	1	319	204.7	301.9	314.2	314.2
20:28:00	2	372	221.3	308	368.8	368.8
20:28:10	2	493	292.9	415.8	478.3	487.9
20:28:20	3	581	359.2	507.8	561.2	584.2
20:28:30	3	665	404.7	584.2	632.8	658.6
20:28:40	3	672	423.8	620.3	658.6	671.9
20:28:50	3	698	449.4	671.9	685.5	699.4
20:29:00	3	790	464.5	685.5	713.5	788.5
20:29:10	3	731	460.6	671.9	713.5	727.9
20:29:20	3	761	467.2	685.5	713.5	727.9
20:29:30	4	746	468.4	685.5	727.9	742.6
20:29:40	4	773	459.6	671.9	713.5	742.6
20:29:50	3	688	434.9	645.6	671.9	685.5
20:30:00	3	904	453	645.6	772.9	804.5
20:30:10	3	704	435.6	632.8	685.5	699.4
20:30:20	4	775	450.1	645.6	713.5	742.6
20:30:30	4	806	446.4	645.6	699.4	727.9
20:30:40	3	724	438.1	632.8	685.5	713.5
20:30:50	3	718	434.4	632.8	671.9	699.4
20:31:00	3	693	433.2	632.8	671.9	671.9
20:31:10	3	690	433	632.8	671.9	671.9
20:31:20	3	697	434.7	632.8	671.9	685.5
20:31:30	4	708	434.3	632.8	685.5	699.4
20:31:40	3	734	440.4	632.8	699.4	713.5
20:31:50	3	742	448.8	658.6	699.4	727.9

20:32:00	4	696	433.9	632.8	671.9	685.5
20:32:10	3	725	435.7	632.8	685.5	699.4
20:32:20	3	714	439.5	645.6	699.4	713.5
20:32:30	3	740	434	632.8	671.9	699.4
20:32:40	4	1434	526.8	645.6	1200.1	1353.1
20:32:50	3	753	439.1	632.8	685.5	742.6
20:33:00	3	716	442.8	645.6	699.4	713.5
20:33:10	4	718	439.3	645.6	685.5	699.4
20:33:20	3	754	447.6	645.6	713.5	742.6
20:33:30	3	751	438.8	632.8	699.4	713.5
20:33:40	3	728	441.2	632.8	699.4	727.9
20:33:50	4	719	435.7	632.8	671.9	699.4
20:34:00	3	727	439.4	645.6	685.5	713.5
20:34:10	3	695	432.9	632.8	658.6	685.5
20:34:20	4	717	438.2	632.8	699.4	713.5
20:34:30	3	686	409.7	632.8	658.6	685.5
Summary	1	1434	433.6	632.8	699.4	742.6

Table 3: Virtual User and Session Details

Time Period	**HTTP Responses**	**Vusers Completed**	**Vusers Created**	**Vusers Failed**
20:27:30	21	7	8	0
20:27:40	36	12	13	0
20:27:50	60	20	20	0
20:28:00	60	20	20	0
20:28:10	66	22	23	0
20:28:20	86	28	30	0

20:28:30	88	29	30	0
20:28:40	94	31	33	0
20:28:50	120	40	39	0
20:29:00	120	40	40	0
20:29:10	124	41	42	0
20:29:20	150	50	50	0
20:29:30	150	50	50	0
20:29:40	150	50	50	0
20:29:50	150	50	50	0
20:30:00	150	50	50	0
20:30:10	150	50	50	0
20:30:20	150	50	50	0
20:30:30	150	50	50	0
20:30:40	150	50	50	0
20:30:50	150	50	50	0
20:31:00	150	50	50	0
20:31:10	150	50	50	0
20:31:20	150	50	50	0
20:31:30	150	50	50	0
20:31:40	150	50	50	0
20:31:50	150	50	50	0
20:32:00	150	50	50	0
20:32:10	150	50	50	0
20:32:20	150	50	50	0
20:32:30	150	50	50	0
20:32:40	150	50	50	0
20:32:50	150	50	50	0

20:33:00	150	50	50	0
20:33:10	150	50	50	0
20:33:20	150	50	50	0
20:33:30	150	50	50	0
20:33:40	150	50	50	0
20:33:50	150	50	50	0
20:34:00	150	50	50	0
20:34:10	150	50	50	0
20:34:20	150	50	50	0
20:34:30	55	20	12	0
Summary	5580	1860	1860	0

Table 4: Session Lengths

Time Period	**Min Session Length (ms)**	**Max Session Length (ms)**	**Mean Session Length (ms)**	**Median Session Length (ms)**	**P95 Session Length (ms)**	**P99 Session Length (ms)**
20:27:30	647.2	714.5	679.4	671.9	713.5	713.5
20:27:40	614.8	684.2	642.6	632.8	671.9	671.9
20:27:50	602.9	646	622.6	620.3	645.6	645.6
20:28:00	598.5	749.9	674.1	671.9	742.6	742.6
20:28:10	754.3	1031.4	893.8	889.1	982.6	1002.4
20:28:20	998.4	1162.1	1077.7	1064.4	1153.1	1153.1
20:28:30	1128.2	1302.9	1213.9	1200.1	1300.1	1300.1
20:28:40	1252.8	1444.6	1277.8	1274.3	1300.1	1353.1
20:28:50	1283.7	1389.6	1360.9	1380.5	1380.5	1380.5
20:29:00	1373.7	1499.6	1408.3	1408.4	1495.5	1495.5
20:29:10	1362.5	1478.8	1391.2	1380.5	1436.8	1465.9

20:29:20	1385.9	1486.8	1415	1408.4	1436.8	1465.9
20:29:30	1382.4	1477.8	1420.9	1408.4	1465.9	1465.9
20:29:40	1281.4	1493.1	1397.5	1408.4	1436.8	1495.5
20:29:50	1277.1	1379.8	1317.3	1326.4	1326.4	1353.1
20:30:00	1293.4	1590.5	1374	1326.4	1495.5	1556.5
20:30:10	1281	1389.9	1316.4	1300.1	1353.1	1380.5
20:30:20	1293.4	1494.1	1362.5	1353.1	1436.8	1465.9
20:30:30	1293.4	1482	1359.4	1353.1	1436.8	1465.9
20:30:40	1288.4	1438.4	1328.2	1326.4	1380.5	1408.4
20:30:50	1286.5	1402.6	1317	1300.1	1353.1	1380.5
20:31:00	1288.2	1385.9	1313.2	1300.1	1353.1	1380.5
20:31:10	1282.4	1353.3	1310.9	1300.1	1326.4	1353.1
20:31:20	1290.4	1392.4	1318.2	1326.4	1353.1	1353.1
20:31:30	1278.9	1382.9	1316.6	1300.1	1380.5	1380.5
20:31:40	1284.4	1443.4	1328.2	1326.4	1408.4	1408.4
20:31:50	1316.7	1449.3	1362.5	1353.1	1408.4	1436.8
20:32:00	1278.6	1456.1	1318.7	1300.1	1353.1	1408.4
20:32:10	1276.5	1410.2	1319.9	1300.1	1380.5	1408.4
20:32:20	1287.6	1399.2	1331.4	1326.4	1380.5	1408.4
20:32:30	1280.7	1414.8	1315.2	1300.1	1353.1	1380.5
20:32:40	1283.1	2499	1598.9	1353.1	2369	2465.6
20:32:50	1281.1	1440.3	1329.9	1326.4	1408.4	1436.8
20:33:00	1295.5	1428.3	1342.8	1353.1	1380.5	1380.5
20:33:10	1281.8	1381.2	1329.4	1326.4	1380.5	1380.5
20:33:20	1293	1446.2	1354	1326.4	1436.8	1436.8
20:33:30	1283.6	1440.4	1332	1326.4	1380.5	1408.4
20:33:40	1280.1	1402	1332.8	1326.4	1380.5	1408.4

20:33:50	1287.8	1402.7	1322.4	1326.4	1353.1	1380.5
20:34:00	1289.2	1409.6	1331	1326.4	1380.5	1380.5
20:34:10	1274.1	1374.4	1310.6	1300.1	1353.1	1353.1
20:34:20	1277.6	1423.3	1328.3	1326.4	1380.5	1408.4
20:34:30	1280.1	1391	1300.8	1300.1	1326.4	1326.4
Summary	598.5	2499	1314.3	1326.4	1436.8	1495.5

Appendix F

Deliverables

Description	Deliverable
Backend GitHub Repository	https://github.com/Archie-Adams/larks/tree/main/backend
Frontend GitHub Repository	https://github.com/Archie-Adams/larks/tree/main/frontend
Demonstration Video	https://youtu.be/q0mxujCOBmc
Autism Detector User Guide	https://github.com/Archie-Adams/larks/wiki/User-Guide-2023

Appendix G
Consent Form

University of Leeds

Consent Form

Title of Project: ______Developing a Digital Health AI Application to Aid Self-Detection of Behavioural Patterns Associated with Autism in Adults_______

Name of Project Student: ________________Ayesha Rahman____________________

Please write either YES or NO in the boxes below to indicate whether or not you agree with the following statements

1	I confirm that I have read and understand the information sheet dated 27 April 2024 *explaining the above project and I have had the opportunity to ask questions about the* project.	YES
2	I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences. In addition, should I not wish to answer any particular question or questions, I am free to decline. Student Contact details: sc21ar@leeds.ac.uk	YES
3	a. I consent to my name being linked to the project materials and to being identified in the reports that result from the project. *or*	
	b. I consent to being identified by my job title and department but not to having my name linked to project materials or the reports that result from the project. *or*	
	c. I consent to take part in the study on the understanding that my responses will be kept strictly confidential, that my name will not be linked with the project materials, and that I will not be identified or identifiable in the reports that result from the project.	YES
4	*I agree for the data collected from me to be used in future research.*	YES
5	I agree to take part in the above project and will inform the project student should my contact details change.	YES

[redacted]	[redacted]	[redacted]
Name of participant	Date	Signature
______Ayesha Rahman______	___27/04/2024___	________________
Name of person taking consent	Date	Signature
To be signed and dated in presence of the participant		
_____Ayesha Rahman______	___27/04/2024___	________________
Project student	Date	Signature
To be signed and dated in presence of the participant		

Re: Interview feedback for Ayesha Rahman Dissertation project

Ayesha Rahman [sc21ar] <sc21ar@leeds.ac.uk>

Today at 01:15

To: [redacted]

Completed on Tuesday 7 May 2024.

Dear [redacted]

I hope this email finds you well.

I would like to express my heartfelt thanks for the time and expertise you shared during the interview regarding my dissertation topic, "Development and Implementation of a Digital Health AI Prototype for Identifying Autism Spectrum Disorder in Adults." Your insights have been invaluable.

Please be assured that your personal details will remain confidential and will not be disclosed in any part of this project.

Thank you once again for your invaluable support.

Kind regards,

Ayesha

Project Information Sheet

Project Title: Development and Implementation of a Digital Health AI Prototype for Identifying Autism Spectrum Disorder in Adults

You are being invited to take part in a student project. Before you decide, it is important for you to understand the aim of the project and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

Project Aim: The project focused on developing an AI-driven application prototype to assist in the self-diagnosis of autism in adults. It has 4 main features – eye movement tracking using model training, Social Interaction Game using NLP, Notes tracking feature to detect autistic characteristics in writing and Questionnaires. The project was done over a year during the student, Ayesha Rahman's, final year project.

Why have I been chosen? You have been recommended by the project supervisor to interview to get insights and feedback for the project-built application for the development of the project. You have been invited to participate in this project to help improve the application's effectiveness.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep (and be asked to sign a consent form) and you can still withdraw at any time. You do not have to give a reason.

What will happen to me if I take part?
If you take part in this project, you will be asked to use the application and give review and feedback for its functions and user interface based on set questions. The application consists of 4 features. For the first feature (eye movement tracking) you need to take a timed 20 second video of you making eye contact with the camera on your device. For the second component of the app (Social interaction game), you will need to give response to set scenarios which will run through a semantic-based NLP. The third component of the application is the note tracking feature which gives the user an area to input their daily routine and feelings. This is then run through a AI model to detect Autistic traits in writing. The fourth component is questionnaires taken from trusted sources. You are not required to complete all the questionnaires if you have time constraints but feedback on the user interface will be asked in the interview.

Will my taking part in this project be kept confidential?
All your personal details and exact responses will be kept confidential. Some specific responses, if used, will be quoted and mentioned in user testing section of the report mentioning no specific details about you. Your name, age, gender, and other such details won't be disclosed but your career role will be mentioned if given consent.

What type of information will be sought from me and why is the collection of this information relevant for achieving the project's objectives? After running through the application, you will have to give responses to the following list of questions:

1. How do you assess the user-friendliness of the application's interface, particularly for users with autism? Are there any elements that might be taken as disruptive?
2. Do you have any suggestions for modifying the colour scheme of the application to better suit its intended user base?
3. The eye movement tracking feature currently uses 20-second video inputs. Do you believe this duration is sufficient to gather meaningful data for detecting autism in adults?
4. Considering that adults who have been misdiagnosed or undiagnosed may often mask their symptoms, do you think this application effectively encourages such individuals to engage and pursue further professional evaluation?
5. Given the limited research on autistic characteristics in written communication, our application attempts to use characteristics from spoken language to analyse diary entries. How effective do you think this approach is, and how could it be improved?
6. In your opinion, does the application provide a supportive platform that could particularly benefit women, who are often underdiagnosed?
7. Are there any additional features you would suggest incorporating into the application to enhance its utility and user experience?
8. For the social interaction game within the application, do you have any recommendations for alternative methods or interface designs that might enhance user comfort and benefit?
9. How do you evaluate the application's measures for ensuring user privacy and data security? Are there any enhancements you would recommend?
10. How well does the application address cultural diversity and accessibility? Are there specific areas where it could be improved to be more inclusive?

What will happen to the results of the project?
The results of this project will be published in a report to be submitted for assessment at the end of the undergraduate module COMP3931 Individual Project in the School of Computing at the University of Leeds.

Contact for further information: sc21ar@leeds.ac.uk , Ayesha Rahman

If you decide to participate in this project, you will be given a copy of this information sheet to keep and may request a copy of your signed consent form at any time. Thank you very much for taking the time to read this information sheet.