

ARI2131 - Course Project

AI-Driven Early Detection System for Dyslexia
and ADHD in Students



**L-Università
ta' Malta**

Name: **Andrea Filiberto Lucas**

Course: Artificial Intelligence (AI)

ID No: 0279704L

1 Abstract

This project primarily focuses on addressing the issue of early detection of learning disabilities in students, with a specific focus on dyslexia and attention deficit hyperactivity disorder (ADHD). The solution conducts a comprehensive analysis of student performance data using AI technology, with the goal of uncovering subtle patterns that could serve as early indicators of these cognitive disorders. To accomplish this goal, the project employs two distinct diagnostic tests: the ADHD test, which examines keyboard and mouse interactions to detect fidgeting behaviors, and the dyslexia test, which assesses how students respond to computer-generated sentences. The findings generated by the AI-driven approach highlights its potential to effectively identify potential learning disabilities, presenting an opportunity for educators and healthcare professionals alike. This tool promises to revolutionize the way these conditions are addressed within the educational landscape by providing an early intervention framework, ultimately improving the prospects and well-being of countless students.

2 Disclaimer

This project is meant solely as a proof of concept and should not be used to diagnose learning disabilities or any medical conditions. This project's AI-driven scripts are experimental and have not been validated by medical authorities. It is critical to note that any conclusions or assessments derived from the use of these scripts should be approached with caution. These scripts are not intended to be a substitute for professional medical evaluations.

3 Introduction

The primary educational challenge addressed within the scope of this project is the timely identification of learning disabilities in students, with a focus on Dyslexia and ADHD. Dyslexia, a learning disability characterized by reading and spelling difficulties, and ADHD, which is known for causing attention-related issues and hyperactivity, both have a significant impact on a student's educational journey and academic performance. Recognizing these disorders at an early stage is critical because it allows for timely interventions that have the potential to significantly improve both educational outcomes and personal development of the effected students.

The significance of addressing this issue extends far beyond the boundaries of academia; it has the potential to reshape the educational paths of many students. In the absence of such early detection systems, students suffering from Dyslexia or ADHD frequently face unnecessary obstacles, resulting in frustration, low self-esteem, and academic setbacks. Educators and parents can proactively implement tailored teaching methodologies and interventions, precisely tailored to accommodate the unique needs of these students, if these learning disabilities are identified during their formative years. This proactive approach not only improves these individuals' educational experiences but also fosters a more inclusive and supportive educational environment for all. As a result, the development of an AI-driven solution aimed at the early detection of these conditions is more than just a technological advancement in the field of education; it is a significant step toward a more equitable and effective educational landscape.

By providing educators, parents, and healthcare professionals with the tools to identify and address these challenges as soon as possible, this project contributes to a future in which every student can unlock their full potential and thrive within a nurturing and inclusive educational environment.

4 Background & Literature Review

4.1 Historical Context - Evolution of Learning Disability Detection

Early detection of learning disabilities (LDs) has been an a long-running challenge in schooling. LDs were previously identified through informal observations made by educators or parents who noticed disparities between a child's academic performance and overall intelligence [1,2]. These subjective methods frequently resulted in delayed diagnoses, making timely implementation of effective interventions difficult [3].

The development of standardised intelligence tests, particularly by Binet and Simon [4], was a watershed moment in the history of identifying such LDs. Intelligence tests provided a more objective method of identifying students who were under performing in comparison to their peers, revealing valuable information about cognitive functioning [5]. However, as educational psychologists and researchers dug deeper into their utility for diagnosing learning disabilities, the inherent limitations of intelligence tests became more clear [6].

One notable limitation was traditional intelligence tests' inability to effectively assess specific learning skills. These tests were created to evaluate general cognitive abilities like verbal and nonverbal reasoning, memory, and processing speed [7]. As a result, they frequently ignored the nuanced difficulties that students with specific learning disabilities, such as dyslexia and ADHD, faced in their academic endeavours. As a result, while intelligence tests provided a useful overall assessment, they lacked the detailed insights required to pinpoint the precise nature of LDs or provide tailored interventions [8].

4.2 Advancements and Challenges in Learning Disability Assessment

The current state of LD assessment technologies and methods is characterised by a variety of approaches, each with its own set of strengths and limitations. Traditional psychological assessment batteries, such as the Wechsler Intelligence Scale for Children (WISC) [9], have long been thought to be the gold standard for evaluating cognitive abilities, including reading abilities. They are, however, time-consuming and costly, rendering them unsuitable for widespread use, particularly in educational settings. On the other hand, emerging technologies such as eye-tracking have the potential to reveal subtle reading patterns and provide valuable insights into learners' difficulties. Nonetheless, issues such as accessibility and eye-tracking technology availability must be addressed [10].

Furthermore, computer-based assessments offer the advantage of real-time data collection, allowing educators and specialists to better monitor students' progress. Their applicability, however, may vary depending on the type of learning disability, indicating the need for a more tailored approach [11]. AI-based solutions have shown promise in analysing large datasets and identifying complex patterns related to learning disabilities [12]. Furthermore, the need for comprehensive cognitive assessments in the identification and treatment of LDs is being called into question, with the need for more efficient and specific methods [7]. Moreover, current computer-based assessments may not be appropriate for all types of learning disabilities [10], as they do not provide personalised assessments. AI has the potential to revolutionise this aspect by enabling highly personalised and adaptive assessments that analyse individual learning profiles and tailor assessments to specific needs [13]. Penalization has the potential to significantly improve intervention effectiveness, benefiting students with learning disabilities in the future [14, 15]. AI also has the potential to improve assessment efficiency, automate data collection, and enable continuous monitoring, further transforming the assessment of learning disabilities in the educational landscape.

5.1 Methodology Sign Posting

5.2 Data Collection and Machine Learning Approach

Subsequently a thorough data cleansing process is carried out on the hypothetical case, addressing any missing values, outliers, and inconsistencies. This stringent procedure ensures that the dataset is ready for analysis.

A combination of well-established Machine Learning (ML) techniques, such as logistic regression, decision trees, a deep learning algorithm, and neural networks, are chosen to address the difficult task of identifying dyslexia and ADHD [16]. Iterative training and hyper parameter tuning are used to fine-tune the models for optimal performance. A comprehensive evaluation that includes metrics such as accuracy, precision, recall, and F-score allow as assessment of the models’ accuracy, in classifying the learning disabilities [17].



Figure 1: AI-Generated (Bing-Chat with GPT-4) icon for the system

5.3 Proof-of-concept design - Dyslexia

"Dyslexia.py" is a proof-of-concept tool that evaluates the likelihood of dyslexia based on user sentence responses. It uses a Natural Language Toolkit (NLTK) for text processing, FuzzyWuzzy for string similarity checks, and ReportLab to generate PDF reports. The script begins by importing libraries and constants, reading sentences from "sentences.txt," and referencing predefined dyslexic letter and word confusions. The central "main()" function manages text processing, user response matching, dyslexia score calculation, and report generation, acting as the analysis command centre.

Based on user responses, the script carefully calculates the dyslexia scores. It examines every word in a response for predefined dyslexic letter matches and word confusions. When a match is found, the word is penalised. The overall dyslexia score is determined by the cumulative penalties for all words in a response. This detailed approach allows for a thorough evaluation of dyslexic characteristics in each user input. The script generates a comprehensive PDF report containing user details, recorded responses, computed dyslexia scores, and a final assessment of dyslexia likelihood after computing dyslexia scores for all responses, providing users with a concise summary of the analysis results to help them understand the dyslexia assessment.

5.4 Proof-of-concept design - ADHD

"Hub.py" and "ADHD.py" are Python scripts designed for user activity monitoring and assessment, with the primary goal of detecting fidgeting and bursts of activity that are indicative of ADHD. The method combines keyboard and mouse event monitoring to gain detailed insights into user actions, as well as the generation of PDF reports that summarise user activity. This solution is made up of two scripts: "Hub.py" handles user interaction and video selection, and "ADHD.py" handles activity monitoring and evaluation. 'webbrowser' and 'subprocess' are used by "Hub.py" for video display and script execution, respectively.

"ADHD.py" makes use of 'pynput' to monitor events and capture keyboard and mouse inputs, as well as 'reportlab' to generate PDF reports. Event listeners categorise user actions in "ADHD.py," and fidgeting detection logic identifies prolonged activity periods. Furthermore, 'matplotlib' visualises user activity over time, allowing for behavioural assessment via burst counting. The video's strategic placement at the top of the screen simulates a classroom-like environment, capturing user engagement, attention, and potential distractions, which is critical for online behaviour assessment.

6 Results and Discussion

6.1 Solution Overview

6.1.1 Dyslexia

The flowchart in figure 2 outlines the steps of a program designed to assess dyslexia. The procedure starts with the execution of a Python script called "Dyslexia.py," which starts the download of necessary NLTK resources, followed by the collection of user information such as name and ID. After that, a random sentence is chosen and the user is asked to rewrite it. Depending on whether the user's response is empty or their rewritten sentence falls below a similarity threshold, the program will either allow for more attempts or calculate a dyslexia score for the attempted sentence. Before the program ends, this score is averaged, a final verdict is determined, and the result is saved as a PDF.

6.1.2 ADHD

Similarly, the following flowchart, figure 3 depicts the ADHD scripts. The procedure begins by running "Hub.py," which displays a large number of subjects for the user to choose from. If the user input is correct, the program displays a video related to the selected subject in a browser. Following that, an external script named "ADHD.py" is run with a two-second delay, recording keyboard and mouse events. If these events occur within 300 seconds, the program continues; otherwise, an ADHD score is calculated. When the score is determined, the script generates a PDF report and graph, both of which are saved locally, and the "ADHD.py" script exits.

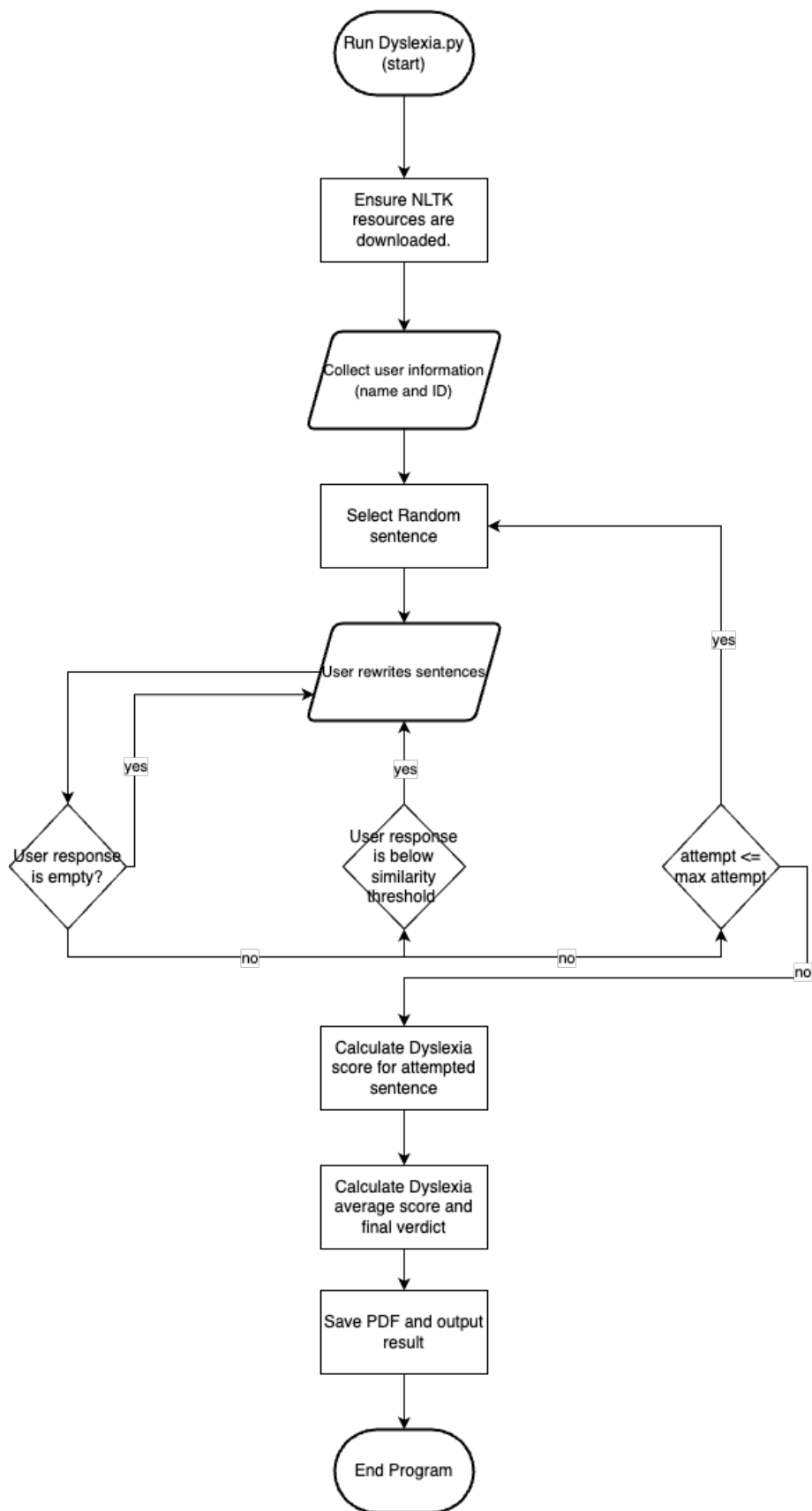


Figure 2: Flowchart of `dyslexia.py`

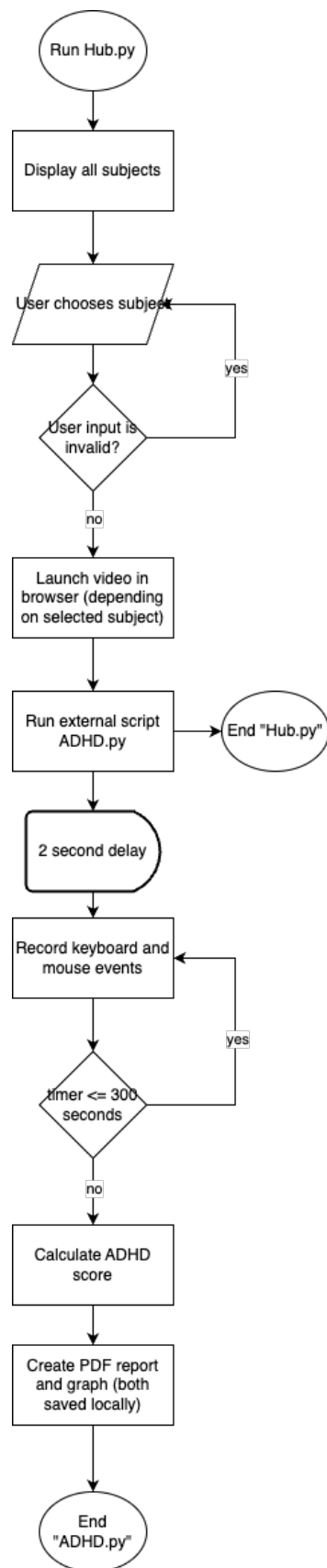


Figure 3: Flowchart of Hub.py & ADHD.py

6.2 Testing and Validation

6.2.1 Dyslexia

The program was run for several times for testing purposes, conducting analyses on two participants one who is a medically certified dyslexic individual and another who is not. Prior to conducting these tests, I obtained the participant's consent to record their data, and will be blurring their full legal name and ID in all documentation to ensure their privacy. Remarkably the program consistently detected dyslexic characteristics in the certified dyslexic participant's input (figure 4) while revealing minimal dyslexic traits in the other non-dyslexic's responses (figure 5). This result demonstrates the program's ability to effectively distinguish between individuals with and without dyslexia, even within the constraints of its current capabilities.

6.2.2 ADHD

To evaluate the program's ability to detect ADHD symptoms, two rounds of testing were conducted. The participant's interest was measured during a video with minimal movement and recorded keyboard strokes in the first round (figure 6). The user simulated ADHD-like behaviour in the second round by increasing keyboard activity and fidgeting intervals (figure 7). The program's response highlighted ADHD symptoms in the generated PDF report, demonstrating its ability to identify ADHD tendencies in users even within its current scope.

Dyslexia Analysis Report

Full Legal Name: [REDACTED]

State ID Number: [REDACTED]

User Responses:

- renewable resources can replenish naturally.
- energy can be kinetic or potential
- the internet of things (iot) connects everyday devices to the internet

Dyslexia Scores:

- Score: 5.400
- Score: 2.500
- Score: 3.643

Average Dyslexia Score: 3.848

Final Verdict: **Likelihood of dyslexia detected.**

Signature: _____




Figure 4: Dyslexia Analysis Report of the medically certified dyslexic participant

Dyslexia Analysis Report

Full Legal Name: [REDACTED]

State ID Number: [REDACTED]

User Responses:

- Don't cry over spoiled milk, move on.
- Test chemistry theories in safe laboratory experiments.
- Bright stars twinkled in the night sky.

Dyslexia Scores:

- Score: 1.714
- Score: 3.810
- Score: 2.143

Average Dyslexia Score: 2.556

Final Verdict: **No significant signs of dyslexia detected.**

Signature: _____




Figure 5: Dyslexia Analysis Report of the non-dyslexic participant

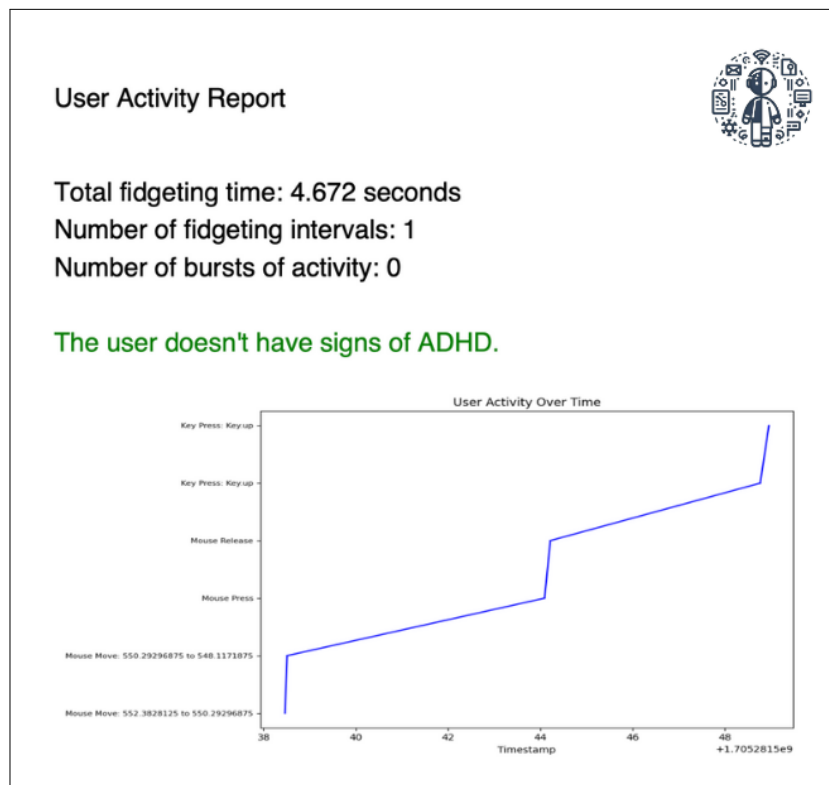


Figure 6: ADHD Analysis Report - negative ADHD result

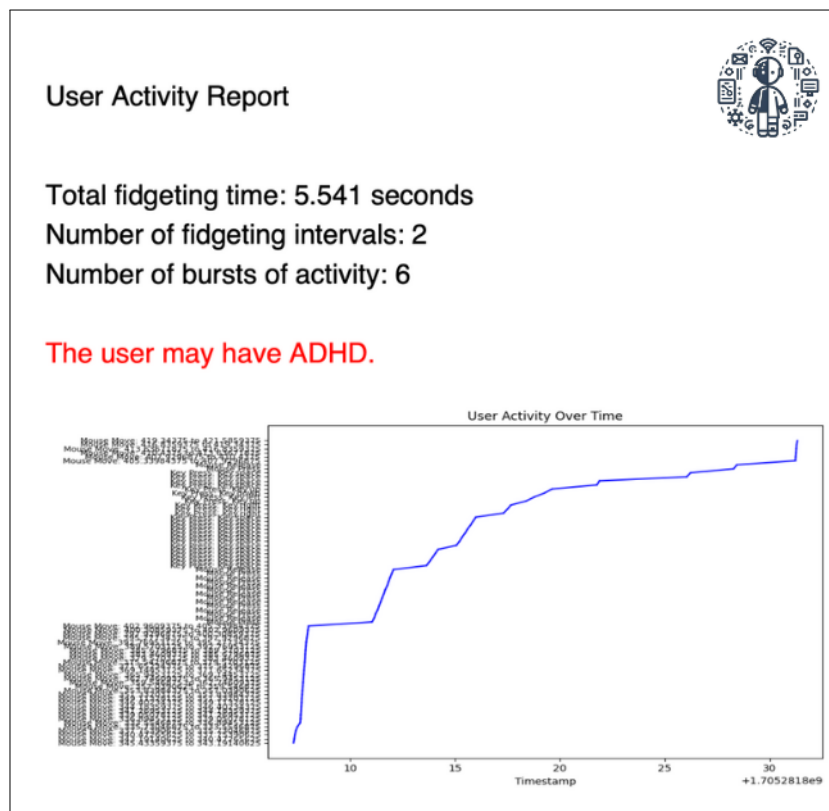


Figure 7: ADHD Analysis Report - positive ADHD result

6.3 Comparative Analysis

Both the AI models for dyslexia and ADHD detection offer significant advantages over traditional non-AI solutions. Traditional assessments for dyslexia frequently necessitate manual evaluations by trained professionals, resulting in delays and increased costs [18]. The AI model, on the other hand, automates the analysis, resulting in a quick, cost-effective, and scalable solution. Its advanced linguistic and pattern recognition capabilities allow it to detect dyslexic features that would otherwise go unnoticed in rule-based or manual evaluations. Similarly, the AI model used in ADHD detection relies on objective data such as keyboard activity and movement patterns to detect ADHD symptoms, reducing the subjectivity associated with traditional methods such as clinical interviews and questionnaires [19]. This objectivity improves assessment accuracy and efficiency, making it a valuable tool for early detection and intervention in both dyslexia and ADHD cases.

7 Ethical Considerations

7.1 Bias and Fairness

Addressing potential biases in the AI model is critical to ensuring that the assessment is fair and accurate to all. Dataset curation, which involves domain experts, is directly related to the project's objectives [20]. Their involvement in data annotation ensures that the data used for training and evaluation is as unbiased as possible [20]. Furthermore, the project's goal of accurate behaviour analysis, particularly in the context of learning disabilities, is consistent with the AI model's predictions for bias being continuously monitored.

7.2 Data Privacy

The project requires the collection of sensitive information about students. The proposed privacy and security measures, such as data encryption, and informed consent mechanisms, are directly related to data protection in the project. Data privacy is essential for maintaining user trust and participation in the study, both of which are critical to the project's success. Furthermore, regular security audits and compliance checks will ensure that the solution complies with data protection regulations such as the EU General Data Protection Regulation (GDPR) [21] or the California Consumer Privacy Act (CCPA) [22].

7.3 Accessibility and Inclusivity

The project's success is dependent on its ability to accurately assess dyslexia and ADHD in a diverse user population. As a result, making the tool accessible and inclusive is directly related to the project's objectives. Design considerations for dyslexic and ADHD users, such as user-friendly instructions and graphical representations, are vital to the project's tool's functionality [23, 24]. Making the tool accessible to users of various abilities aligns with the project's goal of improving behaviour analysis.

8 Limitations & Future Work

8.1 Recognized Limitations

There are several shortcomings in current AI-driven tools for ADHD and dyslexia assessment, especially in their command-line interface (CLI) form. One notable limitation is the lack of a dedicated machine learning algorithm, such as deep learning or neural networks. This omission limits the tools' ability to adapt and learn from data, limiting their ability to improve accuracy. Furthermore, the tools may have difficulty recognising intricate patterns associated with these learning disabilities.

8.2 Recommendations

Incorporating advanced machine learning models, such as neural networks, can greatly improve accuracy and reliability. These models excel at recognising complex patterns and have the potential to improve assessment precision significantly. Additionally, a more diverse and comprehensive datasets would also benefit the tools, addressing the current limitations associated with biased assessments. Data from a variety of demographics, cultural backgrounds, and educational contexts will improve reliability.

Personalisation is also important when assessing learning disabilities. In future iterations, assessments should be tailored to individual needs, learning styles, and prior performance. Personalisation can lead to more accurate assessments and interventions, ultimately improving the tools' overall effectiveness. Adaptation in sentence generation can be implemented by dynamically adjusting sentence complexity and content based on previous user responses. Similarly, when it comes to video subject selection, adaptation can entail analysing the user's preferences and learning patterns in order to recommend subjects that are relevant to their interests and needs.

A future version of the tool could also prioritise the creation of a more graphical user interface (GUI), with an emphasis on accessibility features, making it inclusive and accommodating for users with disabilities. This improved user interface (UI) will ensure that people with varying needs and abilities can effectively access and use the tool.

8.3 Future Potential

The tools have the potential to address a broader range of learning disabilities and disorders, increasing their usefulness for educators and medical professionals. They have the potential to develop into comprehensive platforms that provide not only assessments but also personalised interventions and educational support.

9 Conclusion

In conclusion, the AI-powered Early Detection System for Dyslexia and ADHD in Students is a significant step towards improving the educational landscape. This project aims to provide timely identification of learning disabilities, empowering educators, parents, and healthcare professionals to implement tailored interventions and support by leveraging the power of AI and machine learning. While the current state of the project is just a proof-of-concept, the potential for future iterations is fascinating. This tool has the potential to revolutionise how learning disabilities are identified and managed by addressing limitations, improving personalisation, and improving user interfaces, ultimately leading to a more inclusive and equitable educational system.

10 References

- [1] O'Connor, et al., "Improvement in reading rate under independent and difficult text levels: Influences on word and comprehension skills," *Journal of Educational Psychology*, vol. 102, no. 1, pp. 1–19, 2010. doi: 10.1037/a0017488.
- [2] Peterson, et al., "Developmental dyslexia," *Lancet (London, England)*, vol. 379, no. 9830, pp. 1997–2007, 2012. doi: 10.1016/S0140-6736(12)60198-6.
- [3] J. Salvia, et al., "Assessment in Special and Inclusive Education," 2004.
- [4] "The Development of Intelligence in Children (the Binet-Simon Scale), and the Intelligence of the Feeble-minded," *The Psychological Clinic*, vol. 10, no. 6, p. 175, 1916.
- [5] L. O'Donnell, "The Wechsler Intelligence Scale for Children—Fourth Edition," in *Practitioner's Guide to Assessing Intelligence and Achievement*, J. A. Naglieri and S. Goldstein, Eds. John Wiley & Sons Inc., 2009, pp. 153–190.
- [6] J. Kranzler, et al., "How Do School Psychologists Interpret Intelligence Tests for the Identification of Specific Learning Disabilities?," *Contemporary School Psychology*, Jan. 2020. [Online]. DOI: 10.1007/s40688-020-00274-0.
- [7] Fletcher, et al., "Comprehensive Cognitive Assessments are not Necessary for the Identification and Treatment of Learning Disabilities," *Archives of Clinical Neuropsychology*, vol. 32, no. 1, pp. 2–7, 2017. doi: 10.1093/arclin/acw103.
- [8] J. M. Fletcher, et al., "The validity of discrepancy-based definitions of reading disabilities," *Journal of Learning Disabilities*, vol. 25, no. 9, pp. 555–573, 1992. doi: 10.1177/002221949202500903.
- [9] "Wechsler Intelligence Scale for Children," *Child-Psychologist*, [Online]. Available: <https://www.child-psychologist.com.au/wechsler-intelligence-scale-for-children.html>. [Accessed: 02/01/2024].
- [10] K. Rayner and M. H. Fischer, "Mindless reading revisited: eye movements during reading and scanning are different," *Perception & Psychophysics*, vol. 58, no. 5, pp. 734–747, 1996. doi: 10.3758/bf03213106.
- [11] R. K. Wagner, et al., "Comprehensive Test of Phonological Processing (CTOPP): Examiner's Manual (2nd ed.)," Austin, TX: Pro-Ed, 1999.
- [12] T. S. Poornappriya and R. Gopinath, "Application of Machine Learning Techniques for Improving Learning Disabilities," *Int. J. Electr. Eng. Technol. (IJEET)*, vol. 11, pp. 392–402, 2020.
- [13] T. Hamim, et al., "Survey of Machine Learning Techniques for Student Profile Modeling," in *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 4, pp. 136, 2018. doi: 10.3991/ijet.v16i04.18643.
- [14] L. Major, G. A. Francis, and M. Tsapali, "The effectiveness of technology-supported personalised learning in low- and middle-income countries: A meta-analysis," *British Journal of Educational Technology*, vol. 52, pp. 1935–1964, 2021. doi: 10.1111/bjet.13116.
- [15] A. Haleem, et al., "Understanding the role of digital technologies in education: A review," *Sustainable Operations and Computers*, vol. 3, pp. 275–285, 2022, ISSN 2666-4127, <https://doi.org/10.1016/j.susoc.2022.05.004>.

- [16] S. Dreiseitl, L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," in *Journal of Biomedical Informatics*, vol. 35, no. 5-6, pp. 352-359, 2002. doi: 10.1016/S1532-0464(03)00034-0.
- [17] V. Chang, et al., "An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators," in *Healthcare Analytics*, vol. 2, 2022, p. 100118. doi: 10.1016/j.health.2022.100118.
- [18] N. Mather and D. Schneider, "The Use of Cognitive Tests in the Assessment of Dyslexia," *Journal of Intelligence*, vol. 11, no. 5, p. 79, 2023. [Online]. Available: <https://doi.org/10.3390/jintelligence11050079>.
- [19] T. S. Emser, B. A. Johnston, J. D. Steele, S. Kooij, L. Thorell, and H. Christiansen, "Assessing ADHD symptoms in children and adults: evaluating the role of objective measures," *Behavioral and Brain Functions*, vol. 14, no. 1, p. 11, 2018. [Online]. Available: <https://doi.org/10.1186/s12993-018-0143-x>.
- [20] D. J. Lee and B. Stvilia, "Practices of research data curation in institutional repositories: A qualitative view from repository staff," *PloS One*, vol. 12, no. 3, p. e0173987, 2017. [Online]. Available: <https://doi.org/10.1371/journal.pone.0173987>.
- [21] GDPR.eu. (n.d.). What is GDPR? [Online]. Available: <https://gdpr.eu/what-is-gdpr/>
- [22] California Attorney General's Office. (n.d.). California Consumer Privacy Act (CCPA). [Online]. Available: <https://oag.ca.gov/privacy/ccpa>
- [23] R. Khan, et al., "Proposed user interface design criteria for children with dyslexia," *International Journal of Engineering and Technology*, vol. 7, pp. 5253-5257, 2018. DOI: 10.14419/ijet.v7i4.25496.
- [24] R. U. Khan, et al., "Proposed user interface design criteria for children with dyslexia," *International Journal of Engineering & Technology*, vol. 7, no. 4, pp. 5253-5257, 2018. [Online]. Available: www.sciencepubco.com/index.php/IJET. DOI: 10.14419/ijet.v7i4.25496.

10.1 YouTube Links used in program

- [1] CrashCourse. "Introduction to Biology: Crash Course Biology #1" YouTube.
https://www.youtube.com/watch?v=tZE_fQFK8EY&ab_channel=CrashCourse, 2023
- [2] CrashCourse. "The Nucleus: Crash Course Chemistry #1" YouTube.
https://www.youtube.com/watch?v=FSyAehMdpYI&list=PL8dPuuaLjXtPHzzYuWy6fYEaX9mQQ8oGr&index=2&ab_channel=CrashCourse, 2013
- [3] CrashCourse. "Newton's Laws: Crash Course Physics #5" YouTube.
https://www.youtube.com/watch?v=kKKM8Yu7ds&list=PL8dPuuaLjXtN0ge7yDk_UA0ldZJdhwkoV&index=6&ab_channel=CrashCourse, 2016
- [4] HistoryonMaps. "History of Israel-Palestine Conflict" YouTube.
https://www.youtube.com/watch?v=m19F4IHTVGc&t=5s&ab_channel=HistoryonMaps, 2021
- [5] TED-Ed. "Where do math symbols come from? - John David Walters" YouTube.
https://www.youtube.com/watch?v=eVm063xmnow&ab_channel=TED-Ed, 2017
- [6] HotMess. "Humans and the Environment — Essentials of Environmental Science," YouTube.
https://www.youtube.com/watch?v=RoIpCJwX7-M&ab_channel=HotMess, 2020
- [7] ConcerningReality. "What are the Basic Concepts of Engineering?" YouTube.
https://www.youtube.com/watch?v=u-xjja6mK2k&ab_channel=ConcerningReality, 2020
- [8] CGPGrey. "How AIs, like ChatGPT, Learn" YouTube.
https://www.youtube.com/watch?v=R90Hn5ZF4Uo&ab_channel=CGPGrey, 2017
- [9] AccountingStuff. "ACCOUNTING BASICS: a Guide to (Almost) Everything" YouTube.
https://www.youtube.com/watch?v=yYX4bvQSqbo&ab_channel=AccountingStuff, 2020
- [10] CrashCourse. "Intro to Economics: Crash Course Econ #1" YouTube.
https://www.youtube.com/watch?v=3ez10ADR_gM&ab_channel=CrashCourse, 2015