Universidad de los Andes

Faculty of Engineering
Undergraduate Program in Systems Engineering
Graduation Project
Period 2024-20

EyeNav: A Novel Accessibility-Driven System for Interaction and Automated Test Generation Using Eyetracking and Natural Language Processing

Author: Juan Diego Yepes-Parra
Advisor: Camilo Andrés Escobar-Velásquez

November of 2024
Bogotá, Colombia

Contents

Abstract 3

General Objective 4

Specific Objectives 5

Literature Review 6
- Eye Tracking 6
  - Gaze 6
  - Accuracy and Precision 6
  - Applications 6
- Natural Language Processing 7
  - Applications 7
- Record And Replay Testing 7
  - Gherkin Syntax 7

Related Work 9
- Eyetracking in a chrome extension for behavioral analysis 9
- Eyetracking as an input method 9
- Record and replay testing 9

Methodology 10
- Peripherals 10
  - Tobii Pro Nano Eye Tracker 10
  - Microphone 10
- System Design 11
  - Software Architecture 11
  - Chrome Extension 11
  - HTTP Server 11
  - Websockets 11
  - Testing 11

Results 14
- System 14
- Usability 15
  - System Usabillity Scale 15

Interviews 16
Accesibility 16
Test Scripts Generation 17

Discussion **18**

Future Work **19**

Conclusions **20**

References **21**

1 Abstract

Keywords:

2 General Objective

The primary objective of this project is to leverage eye tracking combined with natural language processing technologies as user input methods for interacting with computers. By combining eye tracking for pointer control and natural language processing for specifying commands, the project seeks to create assistive technology that facilitates computer use. Additionally, the project aims to explore the potential of this interaction model for automating the generation of test scripts in a record-and-replay manner for web applications.

3 Specific Objectives

1. To develop an **eyetracking** system that can be used as an alternative input method for computer interaction.
2. To integrate **natural language processing** capabilities for interpreting user commands and execute the corresponding actions.
3. To develop a framework for **automated test generation** for web applications, in a record-and-replay style.
4. To evaluate the effectiveness of the developed system in terms of **accessibility**, **usability** and accuracy in generating **test scripts**.

4 Literature Review

4.1 Eye Tracking

Eye Tracking refers to "the use of proper techniques and tools for the identification of a subject's gaze direction; in other words, eye tracking allows detecting and recording what users watch, typically, but not necessarily, on a screen" (Cantoni & Porta, 2014). This technique can be measured and applied in different disciplines, which will be explained in the following sections.

Dondi and Porta (2023) point out that there are mainly two different kinds of Eye Trackers:

1. *Remote*, the selected type for the project, which are non-intrusive devices positioned at the bottom of the display.
2. *Wearable*, that are installed in a pair of glasses, for instance, and are able to track outside of digital environments.

4.1.1 Gaze

Gaze is usually associated with a fixed or steady look, however in this field it is used to describe the approximate point on screen where the subject is fixating their eyes. Cantoni and Porta (2014) have explained that to obtain this information in Eye Trackers, gaze direction is obtained during an initial calibration phase, which determines the correspondences between the points observed by the subject on the screen, and the positions of the cornea.

4.1.2 Accuracy and Precision

These two concepts have various applications, like in data science for example, but can be used in this context to understand how the Eye Tracker is performing. Following the manufacturer definitions, accuracy refers to the discrepancy between the true gaze position and the location recorded by the eye tracker, reflecting the systematic error or spatial accuracy of the data. On the other hand, precision indicates the consistency of the eye tracker in capturing the same gaze point across successive samples. It is quantified using the Root Mean Square (RMS) of the recorded points to assess the variation in the data. (Tobii-AB, 2023)

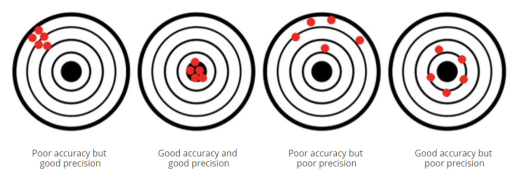


Figure 1: Accuracy and Precision. Source: Tobii-AB (2023)

4.1.3 Applications

- Usability testing

  In this field, eye tracking allows for the analysis of the behavior of the user. This is a process that occurs after they have used the product, which means it does not take an active role in the interaction per se. (Jacob & Karn, 2003)

- Psychology experiments

Many different psychology experiments can be done using eye tracking, for instance, the cognitive load students using a Learning Management System (LMS) has been analyzed using eye tracking, and how it had an effect on their learning outcomes. (Sáiz-Manzanares, Marticorena-Sánchez, Martin Anton, González-Díez, & Carbonero Martín, 2024)

- Input method

  As Jacob and Karn (2003) point out, the application of eye trackers in HCI (Human-Computer Interaction) has long been studied and researched. The approach of using eye tracking as an input method is yet to be applied massively in a commercial environment.

- Artificial Intelligence

  In the rising field of Generative Artificial Intelligence, Taieb-Maimon, Romanovski-Chernik, Last, Litvak, and Elhadad (2024) used eye tracking to perform text summarization, mining data from attention metrics, which outperforms regular text summarization.

4.2 Natural Language Processing

Natural Language Processing (NLP) is a field of artificial intelligence that focuses on the interaction between computers and humans through natural language. The goal of NLP is to enable computers to understand, interpret, and generate human language in a way that is both meaningful and useful. As Kochmar (2022) describes it, "[Natural Language Processing] addresses the various ways in computers can deal with natural—that is human—language"

NLP encompasses several core tasks. These techniques provide the foundation for building more complex systems capable of handling a wide range of applications. For instance, here are some that Kochmar (2022) describes in her book:

- Tokenization

  Breaking text into smaller units such as words, sentences, or subwords to facilitate analysis.

- Syntactic Parsing

  Analyzing sentence structure to identify grammatical relationships between words.

- Semantic Analysis

  Understanding the meaning of text by interpreting context, word sense, and relationships.

- Sentiment Analysis

  Determining the emotional tone or attitude expressed in text.

4.2.1 Applications

4.3 Record And Replay Testing

4.3.1 Gherkin Syntax

All of the testing developed was possible thanks to the Kraken module developed by Ravelo-Méndez, Escobar-Velásquez, and Linares-Vásquez (2023). These tests are made using the Gherkin syntax and WebdriverIO, which enables controlling the browser for automated testing scenarios.

5 Related Work

5.1 Eyetracking in a chrome extension for behavioral analysis

Zelinskyi and Boyko (2024) developed a Chrome Extension that leverages eye tracking combined with session recording to be able to use for behavior analysis.

5.2 Eyetracking as an input method

5.3 Record and replay testing

6 Methodology

6.1 Peripherals

The external hardware required for developing the project will be described in detail.

6.1.1 Tobii Pro Nano Eye Tracker

Figure 2: Tobii Pro Nano Eye Tracker. Source: (Tobii-AB, n.d.)

The Tobii Pro Nano Eye Tracker, shown in figure 2, uses a corneal reflection, dark and bright pupil combination, one-camera system. It supports screen sizes up to 19 inches, introduces a 17 ms system latency. (Tobii-AB, n.d.)

It is characterized by its lightweight and compact form; which makes it easy to carry, and ultimately enabled an easier usability testing process. On the other hand, Tobii Eye Trackers are compatible with the `tobii-research` SDK for python, which allows for controlling the device without the need for a graphical user interface.

This device prompted the idea for the project.

6.1.2 Microphone

The microphone serves as a mediator of interaction, enabling the capture of verbal cues essential for the analysis of words that then get turned into commands thanks to the natural language processing. It also serves as an way of qualitative data acquisition, for recording each participant response for usability testing.

6.2 System Design

Given the system requirements and all of the peripherals it has to handle, a software architectural design was needed, in order to satisfy all the functional and non-functional requirements.

6.2.1 Software Architecture

6.2.2 Chrome Extension

6.2.3 HTTP Server

6.2.4 Websockets

6.2.5 Testing

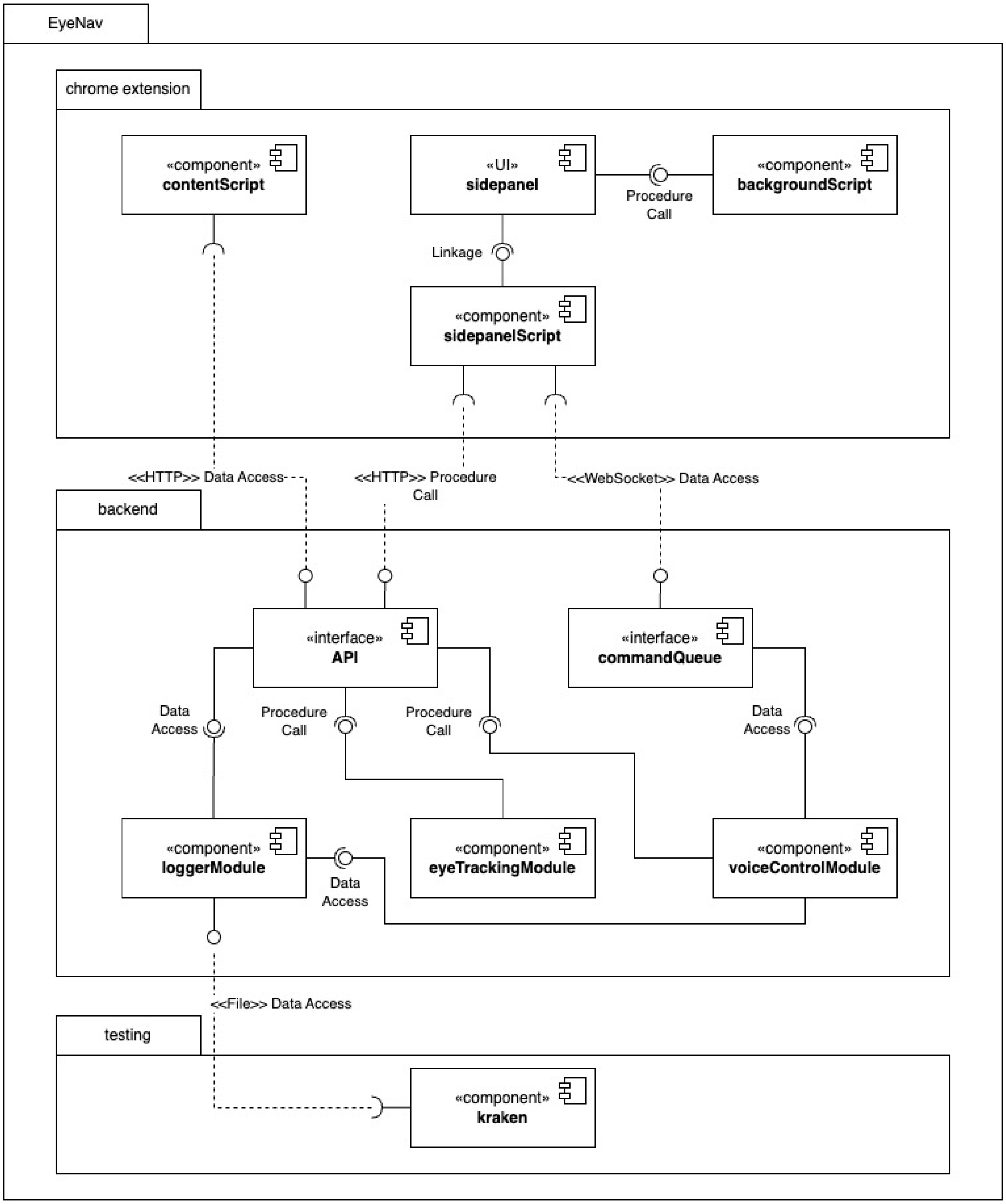


Figure 3: Architecture described in components

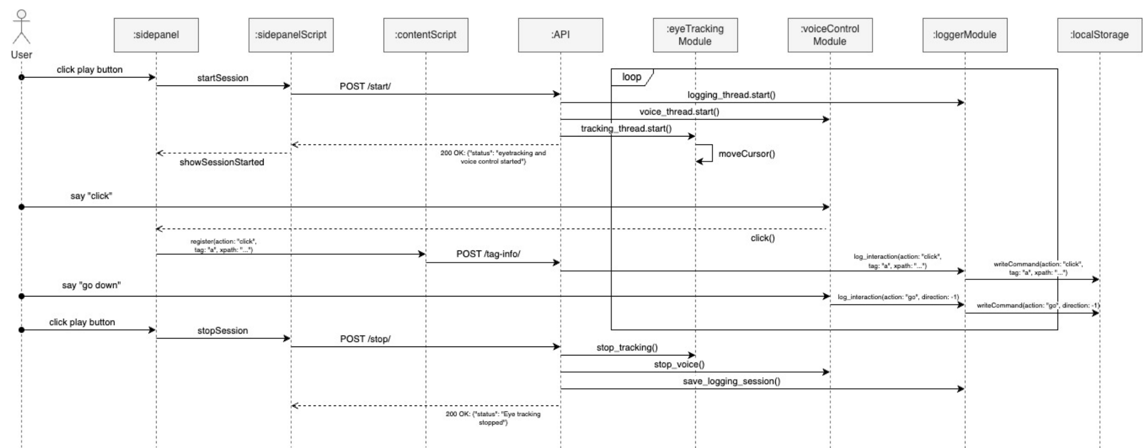


Figure 4: An user clicking and scrolling sequence

7 Results

7.1 System

All of the source code for this project is hosted in a GitHub repository.[1] The repository holds the code for this documentation, the backend module, the chrome extension and the testing module.

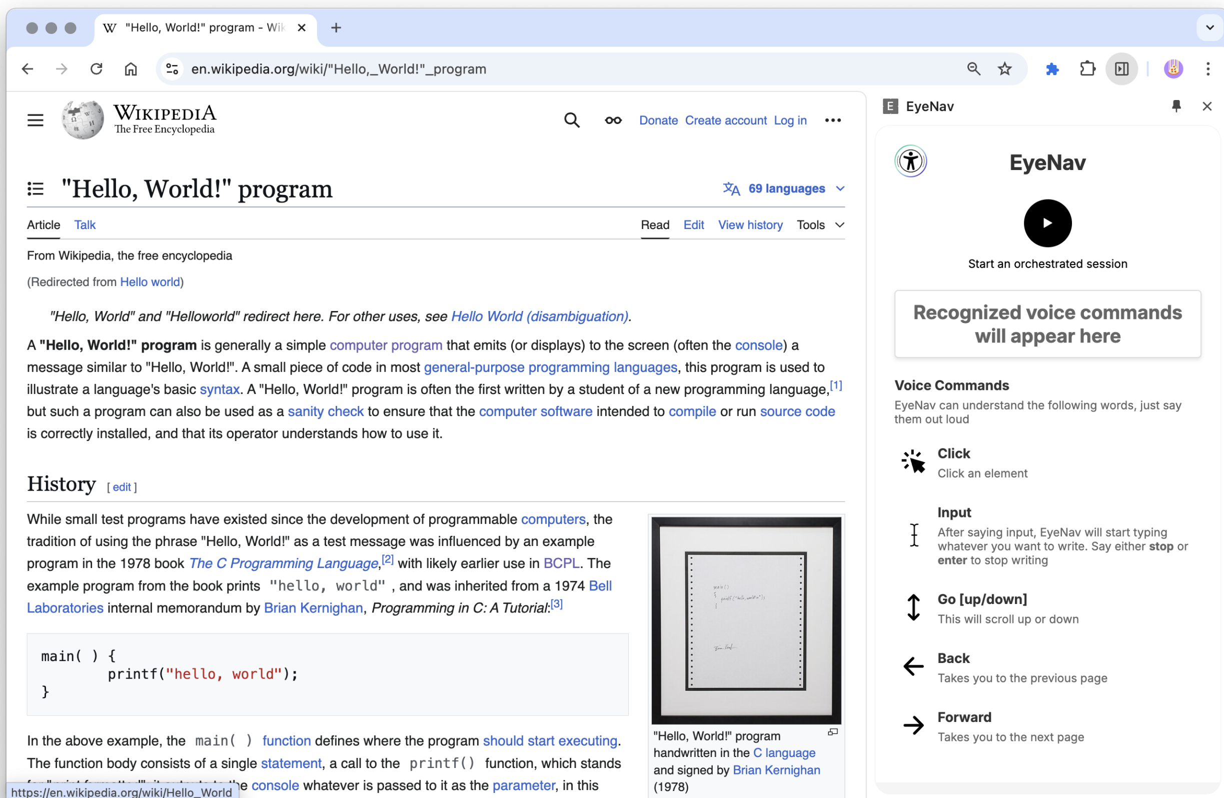
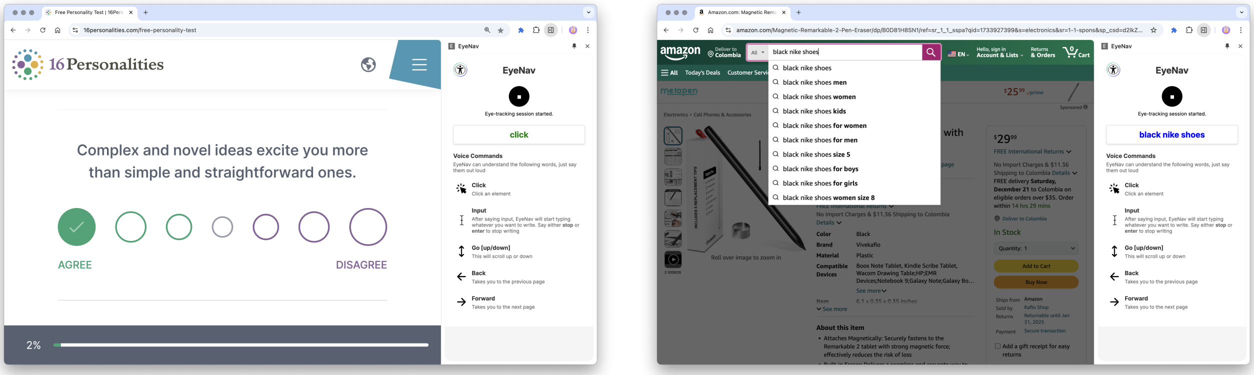


Figure 5: First look of the extension working in Chrome

Figures 5 and 6 display the system in action.

[1]https://github.com/juanyepesp/EyeNav



(a) Clicking an item (b) Inputting text

Figure 6: Other actions that can be done

7.2 Usability

Usability in itself is not an absolute metric, and needs to be defined in particular contexts where it is being applied. In that sense, ways of measuring this non-functional requirement are not straightforward and do not depend on a single method, but are rather cualitative data that need to be analysed in this way.

7.2.1 System Usabillity Scale

The system usability scale (SUS) is useful for understanding broad and general measurements of usability. Proposed by Brooke et al. (1996), this scale is widely used in systems engineering, and consists of 10 questions that the user will assess subjectively, but ultimately can point out how the system in three points:

- effectiveness (the ability of users to complete tasks using the system, and the quality of the output of those tasks)
- efficiency (the level of resource consumed in performing tasks)
- satisfaction (users subjective reactions to using the system). (Brooke et al., 1996)

These are the statements proposed in the scale, where the subject can mark from 1 (strongly disagree) to 5 (strongly agree):

1.	I would like to use this system frequently.	1 2 3 4 5
2.	I found the system unnecessarily complex.	1 2 3 4 5
3.	I thought the system was easy to use.	1 2 3 4 5
4.	I think I would need technical support to use this system.	1 2 3 4 5
5.	The functions of the system were well integrated.	1 2 3 4 5
6.	There was too much inconsistency in the system.	1 2 3 4 5
7.	Most people would learn to use this system very quickly.	1 2 3 4 5
8.	I found the system very cumbersome to use.	1 2 3 4 5
9.	I felt very confident using the system.	1 2 3 4 5
10.	I needed to learn a lot of things before I could get going with this system.	1 2 3 4 5

SUS have scores ranging from 0 to 100. Each individual score can be calculated like formula 1 shows, where each s_n represents each statement

$$2.5\left(20\sum(s_1, s_3, s_5, s_7, s_9) - \sum(s_2, s_4, s_6, s_8, s_{10})\right) \tag{1}$$

Furthermore, in the usability tests conducted, the average SUS was 76.7, with the highest score being 95 and the lowest 55. Indicating that the system was found to be quite usable.

7.2.2 Interviews

7.3 Accesibility

According to the WebAIM (2024), 95.9% of webpages have at least one detectable accessibility error.

7.4 Test Scripts Generation

The tests that the system performs are a set of actions that were recorded when the user was using the system. These actions are written as steps in Gherkin syntax, which are then interpreted and each one corresponds to a Webdriver action the s=testing module can replay. An example test is shown.

```
Feature: Replay of session on Nov 19 at 02:10:49 PM

@user1 @web
Scenario: User interacts with the web page named "Wikipedia, the free encyclopedia"

    Given I navigate to page "https://en.wikipedia.org/wiki/Main_Page"
    And I click on tag with selector a with href "/wiki/Samantha_Harvey_(author)"
    And I click on tag with selector a with href "/wiki/Ditton,_Kent"
    And I click on tag with selector "input" with id "searchInput"
    And I input "hello world"
    And I click on tag with selector a with href "https://en.wikipedia.org/w/index.php?
title=Special%3ASearch&search=%22Hello%2C+World%21%22+program&wprov=acrw1_0"
    And I click on tag with selector a with href "/wiki/Computer_program"
    And I go back
    And I scroll down
```

8 Discussion

Overall, the results indicate success in achieving the research objectives. The integration of the eye tracking system with NLP capabilities proved to be an intuitive and responsive method for computer interaction, as demonstrated by user testing. However, in terms of usability, the system still requires adjustments to better accommodate diverse user bases. This includes improving accessibility for individuals who wear glasses and addressing challenges faced by users speaking different languages or accents, as these groups reported higher levels of frustration when completing tasks.

On the other hand, the testing module provides a flexible and accessible solution for conducting usability tests on any webpage. Its ease of use, accuracy generating the test and adaptability make it a valuable tool for assessing and improving web interfaces in a dynamic environment.

9 Future Work

Considering the scope and impact of the project, the following enhancements and additions can be made to improve the system's accessibility. Furthermore, this study has paved the way for new research areas where the discovered concepts can be applied.

- To improve eye tracking for individuals who wear prescription eyeglasses by using their prescription formula to adjust the system for optimal performance.
- To expand the system's language support and implement internationalization (i18n) features.
- To extend the system capabilities to other environments, such as mobile platforms.
- To include a walkthrough page on how to use the system.
- To incorporate individuals with motor disabilities in future usability testing sessions, as the system has shown potential to be a beneficial solution based on the system usability scale scores and interview feedback.
- To include visual cues for the eyetracking system, to make the system more intuitive
- To leverage record-and-replay test generation as a standalone extension, that can be used with these or other interaction methods.
- To explore the use of these technologies for testing virtual environments, such as those enabled by Meta Quest or Apple Vision Pro mixed reality glasses. This could involve evaluating the responsiveness and usability of these systems, ensuring accessibility features can accommodate a diverse range of users.

10 Conclusions

References

Brooke, J., et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4–7.

Cantoni, V., & Porta, M. (2014). Eye tracking as a computer input and interaction method. In *Proceedings of the 15th international conference on computer systems and technologies* (pp. 1–12).

Dondi, P., & Porta, M. (2023). Gaze-based human–computer interaction for museums and exhibitions: technologies, applications and future perspectives. *Electronics*, *12*(14), 3064.

Jacob, R. J., & Karn, K. S. (2003). Commentary on section 4. eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *The mind's eye*, *2*(3), 573–605.

Kochmar, E. (2022). *Getting started with natural language processing*. Manning.

Ravelo-Méndez, W., Escobar-Velásquez, C., & Linares-Vásquez, M. (2023). Kraken 2.0: A platform-agnostic and cross-device interaction testing tool. *Science of Computer Programming*, *225*, 102897.

Sáiz-Manzanares, M. C., Marticorena-Sánchez, R., Martin Anton, L. J., González-Díez, I., & Carbonero Martín, M. Á. (2024). Using eye tracking technology to analyse cognitive load in multichannel activities in university students. *International Journal of Human–Computer Interaction*, *40*(12), 3263–3281.

Taieb-Maimon, M., Romanovski-Chernik, A., Last, M., Litvak, M., & Elhadad, M. (2024). Mining eye-tracking data for text summarization. *International Journal of Human–Computer Interaction*, *40*(17), 4887–4905.

Tobii-AB. (n.d.). Tobii pro nano: Enter the world of eye tracking research [Computer software manual]. Stockholm, Sweden. Retrieved from `https://tobiipro.com`

Tobii-AB. (2023). *Eye tracker accuracy and precision.* Retrieved from `https://connect.tobii.com/s/article/eye-tracker-accuracy-and-precision`

WebAIM. (2024). *The techreport:webaim-2024 million: The 2024 report on the accessibility of the top 1,000,000 home pages.* (Tech. Rep.). techreport:webaim-2024. Retrieved from `https://techreport:webaim-2024.org/projects/million/`

Zelinskyi, S., & Boyko, Y. (2024). Integrating session recording and eye-tracking: development and evaluation of a chrome extension for user behavior analysis. *Radioelectronic and Computer Systems*, *2024*(3), 38–54.