**AI- Based Image Description Browser Extension for Web Accessibility**

NAME: Justus Ombati Ochego

REG. NO: SC200/0148/2020

COURSE: BSc Computer Science

**A proposal submitted in partial fulfillment of the requirements for the award of Bachelor’s degree in Bsc. Computer Science at the department of Computer Science ,School of Computing and Information Technology, Murang’a University of Technolog**

# DECLARATION

I, Justus Ombati Ochego, hereby declare that the project is the result of my own original work and has not been submitted in any form for another degree or diploma at any university or other institution. All sources of information and assistance used in the project have been acknowledged. Any contribution by others to this project is explicitly mentioned.

I further declare that this work will not be reproduced or submitted by anyone else for academic or non-academic purposes.

NAME…………………………………………

REGISTRATION NUMBER: ………………………………………….

SIGN: ………………………………………….

DATE: …………………………………………

**SUPERVISOR**

NAME: ………………………………………..

SIGN: ………………………………………….

DATE: ………………………………………….

# DEDICATION

This project is dedicated to all those who strive for a more inclusive and accessible digital world. To individuals with visual impairments who inspire us with their resilience and remind us of the importance of creating technology that breaks down barriers. To the developers, designers, and experts who passionately contributed their skills and knowledge, shaping this project into a meaningful endeavor. To the families and friends who provided unwavering support, understanding the significance of fostering inclusivity in the digital landscape. This project is dedicated to a future where technology serves as a bridge, connecting people of all abilities to the vast opportunities of the online world.

# ACKNOWLEDGMENT

This project has been a collaborative journey, and its successful realization owes gratitude to numerous individuals and resources. I extend our heartfelt thanks to Murang’a University support system for their dedication, expertise, and tireless efforts in conceptualizing and bringing this vision to life.

I express our deepest gratitude to the developer, AI specialist, and UX/UI designer, Justus Ombati. whose technical prowess and creativity have shaped the core components of the web browser extension.

Special thanks to the experts in web accessibility, AI, and human-computer interaction who generously shared their knowledge and insights during interviews. Their expertise has enriched the project with a nuanced understanding of industry best practices.

The project would not have been possible without the support of our stakeholders, including users who provided valuable feedback during surveys and testing phases. Their input has been pivotal in ensuring the extension is user-centric and meets diverse needs.

We acknowledge the financial and technical support provided by the parents, friends and well wishers, without which the project's ambitious goals would have been challenging to achieve.

Finally, our gratitude extends to our families, friends, and colleagues for their understanding, encouragement, and unwavering support throughout the project's lifecycle.

This project stands as a testament to the collaborative spirit and collective efforts of all involved parties. Each contribution, big or small, has played a crucial role in bringing this vision to fruition.

# ABSTRACT

This project proposes the development of an innovative web browser extension empowered by artificial intelligence (AI) to enhance web accessibility for individuals with visual impairments. The extension utilizes state-of-the-art image recognition algorithms to automatically analyze and describe images on web pages, providing real-time, contextually relevant descriptions. Through a user-centric approach, the extension allows for customization of image descriptions, catering to individual preferences. The development process follows an agile methodology, emphasizing flexibility and continuous stakeholder collaboration. The project's significance lies in its potential to bridge accessibility gaps, fostering inclusivity in online environments. A comprehensive literature review informs the design choices, and a robust research methodology integrates user feedback and expert insights. The budget and system requirements outline the resources necessary for successful implementation. This endeavor aspires to contribute to the evolving landscape of web accessibility, creating a more inclusive online experience for users with visual impairments.

# TABLE OF CONTENT

Table of Contents

[DECLARATION i](#_Toc153513205)

[DEDICATION ii](#_Toc153513206)

[ACKNOWLEDGMENT iii](#_Toc153513207)

[ABSTRACT iv](#_Toc153513208)

[TABLE OF CONTENT v](#_Toc153513209)

[ACRONYMS AND ABBREVIATIONS viii](#_Toc153513210)

[CHAPTER 1: INTRODUCTION 1](#_Toc153513211)

[1.1 Background Information 1](#_Toc153513212)

[1.2 Problem Statement 2](#_Toc153513213)

[1.3 Objectives 3](#_Toc153513214)

[1.3.1 General Objective 3](#_Toc153513215)

[1.3.2 Specific Objectives 4](#_Toc153513216)

[1.4 Significance of the Study 4](#_Toc153513217)

[1.5 Scope of the Study 5](#_Toc153513218)

[1.6 Limitations 5](#_Toc153513219)

[CHAPTER 2: LITERATURE REVIEW 8](#_Toc153513220)

[2.1 INTRODUCTION 8](#_Toc153513221)

[2.2 EXISTING SYSTEMS 9](#_Toc153513222)

[2.2.1 Google Chrome Extension - Image Captioning by Google AI 9](#_Toc153513223)

[2.2.2 Microsoft Cognitive Services - Image Captioning 10](#_Toc153513224)

[2.2.3 Clarifai Image Captioning 11](#_Toc153513225)

[2.2.5 Proposed System 12](#_Toc153513226)

[2.3 Existing Software Design and Development Tools 14](#_Toc153513227)

[2.4 Literature Review 16](#_Toc153513228)

[2.5 Conclusion 16](#_Toc153513229)

[CHAPTER 3: RESEARCH METHODOLOGY 17](#_Toc153513230)

[3.1 Introduction 17](#_Toc153513231)

[3.2: Data Collection Techniques 17](#_Toc153513232)

[3.2.1 User Surveys: 17](#_Toc153513233)

[3.2.2 Expert Interviews: 18](#_Toc153513234)

[3.2.3 Comparative Analysis of Existing Systems: 19](#_Toc153513235)

[3.2.4 Justification 20](#_Toc153513236)

[3.3: Software Development Techniques 21](#_Toc153513237)

[3.3.1 Agile Development 21](#_Toc153513238)

[3.3.2 Waterfall Model: 22](#_Toc153513239)

[3.3.3 Prototyping 24](#_Toc153513240)

[3.3.4 Justification 25](#_Toc153513241)

[3.4: System Requirements 26](#_Toc153513242)

[3.4.1 Software Requirements 26](#_Toc153513243)

[3.4.2 Hardware Requirements 26](#_Toc153513244)

[3.4.3 Functional Requirements 27](#_Toc153513245)

[3.4.4 Non-Functional Requirements 27](#_Toc153513246)

[3.5 Conclusion 28](#_Toc153513247)

[APPENDIX 29](#_Toc153513248)

[References: 29](#_Toc153513249)

[Budget 32](#_Toc153513250)

[Activity Schedule 33](#_Toc153513251)

**LIST OF FIGURES**

[figure 1:agile methodology illustration [34] 21](#_Toc153479955)

[Figure 2:Waterfall Model [37] 22](#_Toc153479956)

[Figure 3:Prototyping Illustration [39] 24](#_Toc153479957)

**LIST OF TABLES**

[Table 1: Budget 32](#_Toc153483588)

[Table 2:Schedule/ Timeline 33](#_Toc153483589)

# ACRONYMS AND ABBREVIATIONS

AI - Artificial Intelligence

WCAG - Web Content Accessibility Guidelines

CNN - Convolutional Neural Network

RNN - Recurrent Neural Network

API - Application Programming Interface

NLP - Natural Language Processing

ML - Machine Learning

Js - JavaScript

OpenCV - Open Source Computer Vision Library

WWW - World Wide Web

SDK - Software Development Kit

WAI - Web Accessibility Initiative

## CHAPTER 1: INTRODUCTION

### 1.1 Background Information

In today's digital age, the World Wide Web serves as a vast treasure trove of information and entertainment for millions of users across the globe. From educational resources and news articles to social media posts and e-commerce sites, the internet has become an indispensable part of our daily lives.

However, amid this digital abundance, there is a silent challenge faced by a significant portion of the population - those with visual impairments. For these individuals, accessing and comprehending web content, especially images, can be a significant obstacle. Visual elements on web pages, such as infographics, charts, memes, and images, remain elusive unless they are accompanied by alternative text descriptions.[1]

To address this accessibility gap and promote a more inclusive web experience, we present a pioneering project: the development of a web browser extension powered by cutting-edge Artificial Intelligence (AI). This extension is designed to provide detailed, accurate descriptions of images encountered while browsing the internet.

The inspiration for this project arises from a profound desire to bridge the digital accessibility divide. In a world where information is often communicated visually, those without the ability to perceive images are at a considerable disadvantage. This project is not only about enhancing accessibility but also about enriching the online experience for everyone, regardless of their visual capabilities.[2]

Imagine visiting a news website and having images described in detail, enabling users to form a comprehensive understanding of current events. Picture accessing an e-commerce platform where product images are no longer mysteries, but instead, they are accompanied by vivid and informative descriptions. With this web browser extension, our vision is to make these scenarios a reality.

We recognize the immense potential of AI in transforming the accessibility landscape. AI's image recognition capabilities have evolved dramatically, and it is now possible to analyze images and provide textual descriptions that capture the essence of the visual content. Leveraging this technology, our project endeavors to empower individuals with visual impairments and enhance the overall web browsing experience for all users.[3]

The goals of this project are to:

* Develop a web browser extension that can describe images using AI.
* Make the extension easy to use and accessible to all users.
* Evaluate the accuracy and usability of the extension.
* Make the extension available to the public.

How will the project be implemented?

The project will be implemented in the following steps:

1. Research different AI algorithms for image description and choose an algorithm to use in the extension.
2. Develop the extension and integrate it with the chosen AI algorithm.
3. Test the extension and evaluate its accuracy and usability.
4. Make the extension available to the public.

Who will benefit from the project?

The project will benefit people with disabilities, such as blindness or low vision, as well as people with dyslexia. In addition, the project could also benefit social media users and creators.

### 1.2 Problem Statement

The World Wide Web is a vast repository of information and visual content, with images playing a pivotal role in conveying ideas, information, and emotions. However, this digital treasure trove remains significantly inaccessible to a sizable segment of the global population—those with visual impairments. The fundamental challenge lies in the fact that, in the absence of alternative text descriptions, individuals with visual disabilities are denied access to the information and context embedded within these images. This disparity in web accessibility not only hinders the ability of these users to fully participate in the digital age but also diminishes the richness of the online experience for everyone.[4]

As of now, web content creators often neglect or inadequately provide alternative text descriptions for images, rendering the images on web pages invisible and inscrutable to screen readers used by visually impaired individuals. This exclusion extends to educational materials, news articles, e-commerce platforms, social media posts, and more, creating a digital divide where visual information is the key to understanding and engagement.

The problem is two-fold: first, images serve as a significant component of web content, and second, existing solutions for generating alternative text descriptions often rely on manual input, which is inconsistent, time-consuming, and frequently neglected. The need for a robust, automated solution that can accurately and comprehensively describe images encountered while browsing the web is evident. An innovative approach that leverages Artificial Intelligence (AI) and integrates seamlessly into web browsers can revolutionize web accessibility, enrich the online experience for individuals with visual impairments, and enhance the usability of the internet for everyone. (WCAG 2.2) [5]

To address this issue, our project seeks to develop a web browser extension that harnesses the power of AI to automatically describe images on web pages, providing a detailed and meaningful textual representation of visual content. By solving this problem, we aim to make the web more inclusive, empowering individuals with visual impairments and ensuring that web content is universally accessible and informative. This project represents a vital step toward a more equitable and inclusive digital future, where everyone, regardless of their visual capabilities, can fully participate in the online world.

### 1.3 Objectives

#### 1.3.1 General Objective

To develop a web browser extension that employs AI technology to automatically describe images on web pages, enhancing web accessibility and inclusivity for users with visual impairments, while improving the overall browsing experience for all users.

#### 1.3.2 Specific Objectives

* To Define project scope, objectives, and requirements to ensure a clear understanding of the extension's purpose and functionality, as well as the specific needs of users with visual impairments.
* To conduct a thorough analysis of existing web accessibility tools and technologies, identify key user requirements, and evaluate AI algorithms for image description to determine the most suitable approach.
* To create a detailed design plan for the web browser extension, including user interfaces, data flow, and integration with web browsers, ensuring a user-friendly and seamless experience for all users.
* To develop the web browser extension, integrating AI image description algorithms and testing their accuracy in real-time image recognition and description. Implement efficient coding practices and ensure compatibility with major web browsers.
* To rigorously test the web browser extension to ensure its reliability, accuracy, and accessibility compliance. This includes functional, usability, and accessibility testing to validate the extension's effectiveness in describing images on diverse web pages.
* To deploy the web browser extension to a range of web browser platforms, promote its availability to the target audience, and provide ongoing maintenance and support to ensure its continued functionality, security, and relevance.

### 1.4 Significance of the Study

The significance of this study lies in its capacity to empower individuals with visual impairments, providing them with inclusive access to the visual content of the World Wide Web, thus fostering digital inclusivity and creating a more equitable and enriched online experience for users of all ablities. By leveraging cutting-edge AI technology, this project not only bridges the accessibility gap but also contributes to the ongoing evolution of web technology, reinforcing the web as a space that promotes diversity, equality, and universal access to information and culture.[6]

### 1.5 Scope of the Study

The scope of this study and project encompasses the development of a web browser extension that utilizes advanced Artificial Intelligence (AI) technology to automatically describe images found on web pages. This extension will be designed to be compatible with major web browsers, enabling it to function seamlessly across various online platforms and websites. The primary focus of the project is to address the accessibility and inclusivity challenges faced by individuals with visual impairments, ensuring that they can access, comprehend, and engage with visual content on the internet. The scope includes the development of an AI-based image recognition and description system, integration with web browsers, and user interface design that caters to both sighted and visually impaired users. A(Xiv preprint arXiv:1409.1556 (2014)). [7]

Additionally, the project will involve the testing and validation of the extension's functionality, usability, and accessibility compliance to guarantee that it effectively and accurately describes images across diverse web content. The study will encompass ongoing maintenance and support to keep the extension up-to-date and in line with evolving web technologies. While the core focus is on accessibility for individuals with visual impairments, the extension's benefits extend to all users, enhancing their web browsing experience by providing insightful and informative image descriptions. The scope of this study, therefore, extends to fostering a more inclusive, diverse, and enriching online environment where information and culture are accessible to all, regardless of their visual capabilities.

### 1.6 Limitations

Here are some potential limitations of the study of developing an AI powered browser extension:[8]

1. **Image Recognition Accuracy:** The accuracy of AI image recognition algorithms is a critical factor in the success of the project. While AI has made significant advancements in this area, it may still have limitations in accurately describing complex or abstract images. Variability in image context, quality, and content may pose challenges, and the extension's effectiveness may be contingent on the quality of the underlying AI model.
2. **Web Page Compatibility:** The extension's compatibility with all web pages and platforms may not be absolute. Some web pages may use unconventional coding or structures that could affect the extension's ability to access and describe images. Achieving universal compatibility may require continuous updates and adjustments.
3. **User Engagement:** The success of the extension relies on users adopting and actively using it. The project's limitation is that it cannot guarantee widespread adoption, and user engagement may vary depending on factors such as awareness, user preferences, and the need for image descriptions.
4. **Processing Resource Requirements:** The AI-based image recognition and description process may demand substantial processing resources, potentially affecting the extension's performance and efficiency. This limitation may impact the extension's usability on older devices or those with limited processing power.
5. **Language and Multilingual Support:** The project may face limitations in providing image descriptions in multiple languages. The AI models used may excel in certain languages but may not perform as accurately in others. Addressing this limitation might require additional language-specific training and resources.
6. **Privacy and Data Usage:** To function effectively, the extension may need to access and process data from web pages, raising privacy concerns. Striking a balance between providing image descriptions and respecting user privacy may present a limitation, necessitating robust data privacy measures. a( Amodei, Dario, et al.)[9]
7. **Continuous Maintenance:** Ensuring the extension remains up-to-date and secure may be an ongoing challenge. Web technologies, browser updates, and evolving AI models will require continuous maintenance and updates, which could be resource-intensive.

## CHAPTER 2: LITERATURE REVIEW

### 2.1 INTRODUCTION

This research comprises of one of the most important aspect-the literature review that forms the basis of the development of its conceptual framework. Therefore, this is a chapter that provides a wide review of published articles on web accessibility, image recognition and use of Artificial Intelligence in meeting the need of those who are visually impaired. Accessibility through internet has been identified over time as one of the most vital issues, where people need to have an opportunity to view the same information and enjoy the same services regardless of their individual abilities. Web Accessibility for People with Visual Impairments: A Survey and Analysis, the title chapter begins with a discussion of problems currently affecting people who experience visual impairment, as it relates to being able to read or view graphic material available on the modern website, which leads them to feel frustrated In order to provide the relevant background for the creation of an artificial intelligence (AI)-enabled web browser extension with image enhancement functionality, this chapter scrutinizes various reports and industry practices.[10]  
  
The following literature review then traces the history of AI development, particularly as it relates to image recognition and description. Recent enhancement of deep learning as well as neural network models has enabled the AI system to understand and make a statement about the object featured in an image. Such discussion considers the development of the image recognition algorithms, outlining application areas, as well as the shortcomings giving pre-requisite knowledge on the technological background of our proposed web browser extension. The objective for this chapter is to highlight on gaps in the current methods by thoroughly scrutinizing existing research and thereby determine the specific contribution and significance of the proposed project. In the following chapters, the synthesis of literature will lead to designing and deployment of the AI algorithms in this regard so that the developed extension would be up-to-date concerning both web accessibility and artificial intelligence.[11]

### 2.2 EXISTING SYSTEMS

#### 2.2.1 Google Chrome Extension - Image Captioning by Google AI

Google Chrome Extension - Image Captioning by Google AI utilizes a sophisticated deep learning architecture to analyze images and generate descriptive captions. The heart of this architecture is a convolutional neural network (CNN), which is a type of artificial neural network that excels at extracting visual features from images. The CNN in this extension is trained on a massive dataset of images and their corresponding captions, allowing it to learn the intricate relationships between visual patterns and language. [12]

When an image is presented to the extension, the CNN first extracts a comprehensive representation of the image's visual content. This representation captures the essential features of the image, such as the presence of objects, their shapes, colors, and spatial relationships. The extracted visual representation is then passed to a recurrent neural network (RNN), which specializes in processing sequential data like language. The RNN takes as input the sequence of visual features and generates a corresponding sequence of words, forming a coherent and descriptive caption that accurately reflects the image's content.

The extension seamlessly integrates with the Google Chrome browser, making it easy to activate and use. Simply clicking on an image within a web page triggers the extension, and the generated caption is promptly displayed alongside the image. This user-friendly interface provides a convenient way to gain insights into the visual content of web pages.

**Advantages**:

* Accurate image descriptions: The CNN model is trained on a massive dataset of images and their corresponding captions, resulting in highly accurate image descriptions.
* Ease of use: The extension is easy to install and use, requiring minimal user input. (Balakrishnan & Narasimhan, 2010)[13]
* Integration with Google Chrome: The extension seamlessly integrates with Google Chrome, providing a convenient way to describe images on the web.

**Disadvantages**:

* Limited to English language: The extension currently only supports image descriptions in English.
* Limited to static images: The extension cannot describe dynamic images or videos.
* Potential for bias: The CNN model may exhibit biases based on the data it was trained on.[14]

#### 2.2.2 Microsoft Cognitive Services - Image Captioning

Microsoft Cognitive Services - Image Captioning offers a powerful image captioning service that leverages the combined strengths of convolutional neural networks (CNNs) and recurrent neural networks (RNNs). CNNs, with their ability to extract visual features from images, provide the foundation for understanding the visual content. Meanwhile, RNNs, with their ability to process sequential data like language, take the extracted features and generate corresponding captions.[15]

The service's image captioning process begins with a CNN analyzing the input image. The CNN extracts a detailed representation of the image's visual content, capturing the presence of objects, their shapes, colors, and spatial relationships. This rich visual representation is then passed to an RNN, which takes into account the context of previous words to produce a coherent and natural-sounding caption. The RNN employs a sophisticated language model that incorporates grammatical rules and semantic knowledge to ensure the generated captions are accurate and fluent.

Microsoft Cognitive Services - Image Captioning is accessible through an API, allowing it to be integrated into various applications, including web extensions. This flexibility makes the service a versatile tool for enhancing the accessibility and usability of web content.

**Advantages**:

* Accurate image descriptions: The combination of CNNs and RNNs provides high-quality image descriptions.
* Support for multiple languages: The service supports image descriptions in multiple languages, including English, French, German, Spanish, and Chinese.
* Integration with various applications: The API-based nature of the service allows integration into various applications, including web extensions. [16]

**Disadvantages**:

* Requires an API key: Using the service requires obtaining an API key, which may involve costs.
* Potential for bias: The CNN and RNN models may exhibit biases based on the data they were trained on. [17]

#### 2.2.3 Clarifai Image Captioning

Clarifai Image Captioning combines the power of deep learning and natural language processing (NLP) to generate accurate and creative image descriptions. Deep learning models excel at extracting meaningful features from images, while NLP techniques are adept at processing and generating human language. By combining these two approaches, Clarifai Image Captioning produces captions that are not only factually accurate but also engaging and evocative.

The process begins with a deep learning model analyzing the input image. The model identifies and extracts relevant features, such as the presence of objects, their attributes, and their relationships. These features are then used to generate a set of candidate captions.

Next, NLP techniques are employed to evaluate the generated captions. The NLP models assess the captions' relevance to the image, their coherence and grammatical correctness, and their overall creativity. The most highly rated captions are selected as the final output, ensuring that the generated descriptions are not only accurate but also engaging and insightful.

Clarifai Image Captioning is available through an API, enabling integration into various applications, including web extensions. This integration allows users to seamlessly generate descriptive captions for images within their web browsing experience. [18]

**Advantages**:

* Accurate and creative image descriptions: The combination of deep learning and NLP provides accurate and creative image descriptions.
* Support for multiple languages: The service supports image descriptions in multiple languages, including English, French, German, Spanish, and Chinese.
* Integration with various applications: The API-based nature of the service allows integration into various applications, including web extensions. [19]

**Disadvantages**:

* Requires an API key: Using the service requires obtaining an API key, which may involve costs.
* Potential for bias: The deep learning and NLP models may exhibit biases based on the data they were trained on.

2.2.5 Proposed System  
This paper discusses a system that will introduce web based browser extension with state of the art intelligence AI algorithms used automatically in describing images. This extension strikes a compromise between the advantages of automation and the flexibility of manual annotation compared with existing systems. Its main function revolves around real time image processing on web sites providing semantically rich descriptions that improve the user experience in a web accessible environment.

**Key Features:**  
  
1. AI-Powered Image Recognition: It uses modern AI image recognition systems that can autonomously analyze and describe the contents of pictures viewed on the web.  
  
2. Real-Time Processing: This extension operates in real-time unlike the previous manual annotation systems. Therefore, users are able to obtain instant and correct image descriptions with no additional user input required.  
  
3. Browser Compatibility: It will plug into top-line browsers, and it will be just one solution to accommodate all web-browsing needs.  
  
4. Customization Options: The advantage is that users decide on how detailed or general the image description should be thus having the opportunity to suit their personal needs of information. The need to offer custom solutions for the purposes of accessible is also considered with this feature.  
  
5. Offline Mode: We will develop an offline mode that would help reduce the dependency of the extension in use on the internet connectivity. The extension will continue offering the image description in situations of limited internet access using information stored in data storage.  
  
6. Resource Optimization: Optimization is used as an extension technique that enables effective performance on different devices such as high-end computers and lower end ones.  
  
7. Privacy Measures: Thus, the system is designed with strong mechanisms for safeguarding user information and ensuring user privacy. Whenever achievable, local image analysis is performed for minimizing external server dependencies.

This proposed system aims to represent a comprehensive solution that combines the strengths of existing approaches, leveraging AI for automation while offering users a degree of control and customization. Through its innovative features, the extension seeks to significantly improve the accessibility of web content for individuals with visual impairments, fostering a more inclusive online environment.

### 2.3 Existing Software Design and Development Tools

In the realm of software design and development, selecting appropriate languages, frameworks, and libraries is crucial for the success of any project. The proposed web browser extension, aimed at enhancing image accessibility through AI, benefits from a thoughtful choice of tools. The selection encompasses programming languages, frameworks, and libraries that collectively contribute to efficient development, seamless integration, and optimal performance. [20]

**Programming Languages:**

1. JavaScript: As a fundamental language for web development, JavaScript is essential for creating dynamic and interactive user interfaces. It is the backbone for many browser extensions and provides the necessary functionality for integrating AI algorithms seamlessly into web pages.

2. Python: Widely recognized for its versatility and rich ecosystem of libraries, Python is well-suited for implementing AI and machine learning components. Python can be employed for training and utilizing AI models for image recognition.

**Frameworks**:

1. TensorFlow.js: An open-source machine learning framework developed by Google, TensorFlow.js allows the execution of machine learning models directly in the browser. Its JavaScript library facilitates the integration of pre-trained models for image recognition into the web extension. [21]

2. React.js: For building user interfaces with a modular and component-based approach, React.js is a popular choice. Its efficiency in managing UI components aligns well with the dynamic nature of a browser extension.

**Libraries**:

1. WebExtension API: This API provides the necessary tools for developing browser extensions compatible with major browsers like Chrome, Firefox, and Edge. It allows for interactions with browser functionality and content scripts for webpage interaction.

2. OpenCV (Open-Source Computer Vision Library): A powerful library for computer vision tasks, OpenCV can be utilized for image processing and analysis. Its capabilities can enhance the accuracy of image recognition algorithms within the extension. [22]

3. D3.js (Data-Driven Documents): If data visualization becomes a significant aspect, D3.js can be valuable. It allows for the creation of dynamic and interactive data visualizations that can be integrated into the extension's user interface.

**Development Tools:**

1. Visual Studio Code: A lightweight, yet powerful, source code editor that provides excellent support for JavaScript, Python, and various web technologies. Its extensibility allows the integration of tools specific to the project's needs.

2. Git and GitHub: Essential for version control and collaborative development, Git in conjunction with platforms like GitHub enables efficient code management, collaboration, and issue tracking.

By leveraging this set of languages, frameworks, and libraries, the development process is streamlined, allowing for the integration of AI capabilities into the web extension effectively. The chosen tools provide a balance between efficiency, compatibility, and performance, ensuring the successful realization of the proposed system.

### 2.4 Literature Review

This comprehensive literature review serves as a foundational pillar for the proposed web browser extension project. By systematically examining existing systems, we gain valuable insights into the strengths, limitations, and innovations within the domain of web accessibility and AI-powered image recognition. Understanding the workings of screen readers, browser extensions, and AI models in current applications informs our design choices, helping us distill the most effective features while addressing potential challenges. The justification lies in the meticulous analysis of each existing system, providing a knowledge base that guides the proposed system's development. This literature review not only identifies gaps in current solutions but also positions the project within the broader context of advancements in web technology and accessibility, ensuring its relevance and contribution to the field. [23]

### 2.5 Conclusion

In conclusion, the literature review has provided a comprehensive exploration of existing systems related to web accessibility and image recognition. By synthesizing knowledge from screen readers, manual annotation browser extensions, and AI-powered solutions, we have gained a nuanced understanding of the current landscape. This knowledge is pivotal for informing the design and development of the proposed web browser extension. As we embark on this project, the insights garnered from the literature review will guide our efforts to create a system that not only addresses the limitations of existing solutions but also integrates seamlessly into the evolving landscape of web accessibility, thereby contributing to a more inclusive online experience.

## CHAPTER 3: RESEARCH METHODOLOGY

### 3.1 Introduction

In this research, the third chapter details how objective of the proposal was achieved using different methods. This section makes it easier to outline or write down a simple step of the project with regard to its design, production and evaluation. It goes without saying that the choice of a particular methodology in this case has its significance since developing a web browser extension implies accurate and reliable results in terms of innovation. The next sections will explore in detail the research process; providing details about the how the proposed AI-powered extension was designed and implemented. Specifically, the methodology entails choosing the appropriate tools, defining steps in developing a program, and evaluation metrics for assessing whether or not the extension has met its objectives. Following a proper methodology, this study intends to significantly contribute in areas of web Accessibility and Artificial Intelligence by making the report reliable and replicable at all times. (Creswell, 2018; Leedy & Ormrod, 2001). [24]

### 3.2: Data Collection Techniques

#### 3.2.1 User Surveys:

User surveys represent a valuable technique for gathering qualitative and quantitative insights from potential users regarding their expectations, preferences, and challenges related to web accessibility and image description. Surveys can be distributed to a diverse group of users, including those with visual impairments and sighted users, to capture a holistic perspective on the proposed web browser extension. Questions may focus on the desired features, user interface preferences, and the perceived effectiveness of existing accessibility tools.

There are many different types of user surveys, but some of the most common include:

**Customer satisfaction surveys:** These surveys measure how satisfied users are with a product or service.

**Usability surveys:** These surveys measure how easy a product or service is to use.

**User experience surveys:** These surveys measure the overall experience that users have with a product or service. [25]

**Advantages**

* Diverse User Input: Surveys allow for the collection of feedback from a broad user base, ensuring that the extension is designed to cater to various user needs and preferences.
* Quantifiable Data: The structured format of surveys enables the generation of quantifiable data, facilitating statistical analysis that can inform key design decisions.
* Early User Engagement: By involving users in the design phase, potential pitfalls and user requirements can be identified early, guiding the development process effectively.[26]

**Disadvantages**

* Limited Depth: Surveys may provide surface-level insights, and nuanced user experiences may not be fully captured, necessitating supplementary methods for in-depth understanding.
* Potential Bias: The respondents' perspectives might be influenced by their familiarity or lack thereof with existing accessibility tools, potentially introducing bias into the collected data.
* Dependence on User Availability: Relying on user participation may lead to varying response rates, potentially affecting the comprehensiveness of the collected data. [27]

#### 3.2.2 Expert Interviews:

Expert interviews involve engaging with professionals in the fields of web accessibility, AI, and human-computer interaction. Through structured interviews, insights can be gained into best practices, potential challenges, and innovative approaches. Experts can provide critical perspectives on the technical aspects of implementing AI for image description, ensuring that the proposed extension aligns with industry standards and advancements.[28]

**Advantages:**

* In-Depth Insights: Interviews facilitate detailed discussions, allowing experts to share nuanced insights that may not be captured through other data collection methods.
* Validation of Technical Approaches: Experts can validate the feasibility and effectiveness of proposed technical solutions, ensuring that the project aligns with current industry standards.
* Identification of Potential Pitfalls: Experts can anticipate and highlight potential challenges in the development phase, enabling preemptive problem-solving. [29]

**Disadvantages:**

* Limited User Perspective: While experts provide invaluable technical insights, their views may not fully represent the diverse perspectives and needs of end-users, necessitating a balance with user-centric methods.
* Resource Intensive: Scheduling and conducting expert interviews may be resource-intensive, requiring coordination and time commitments from both the researchers and the experts.
* Potential for Expert Bias: The experts' perspectives may be influenced by their specific expertise, potentially introducing bias into their recommendations. [30]

#### 3.2.3 Comparative Analysis of Existing Systems:

A comparative analysis involves a thorough examination of existing web accessibility tools, image recognition algorithms, and browser extensions. This technique allows for the identification of strengths and weaknesses in current solutions, guiding the design choices for the proposed extension. By dissecting the underlying technologies and methodologies of existing systems, the research can pinpoint areas for improvement and innovation. [31]

**Advantages:**

* Informed Design Choices: Analyzing existing systems provides a foundation for making informed design decisions, drawing on successful elements while addressing shortcomings.
* Identification of Innovation Opportunities: Comparative analysis helps identify gaps in current solutions, presenting opportunities for introducing novel features or enhancements.
* Risk Mitigation: Understanding the strengths and weaknesses of existing systems allows for proactive mitigation of potential challenges and pitfalls in the design and development phases.

**Disadvantages:**

* Dynamic Technological Landscape: The rapid evolution of technology may result in a lag between the analysis phase and the actual development, requiring continuous monitoring to stay abreast of the latest advancements.
* Limited Customization Insights: Existing systems may lack certain customization features or may not fully cater to the diverse needs of users, necessitating additional user-centric research.
* Dependency on Publicly Available Information: Access to proprietary information about certain systems may be limited, potentially constraining the depth of the comparative analysis. [32]

#### 3.2.4 Justification

Careful matching of the selected tools for data collection to the initial stages of a project, that is, the design & analysis. The design team can then use these user surveys to have an idea of what the users expect regarding usability as well as their perceptions with regards various features of the extension. Validation of technical approaches, including comparison in accordance with the established requirements, and conformity of the design solution with industry standards are achieved through expert interviews, which contributes to the reliability of design solution. Comparing existing systems creates a baseline on which informed design decisions can be made to ensure that the project avoids the pitfalls and offers opportunities to innovate.

Combined, these techniques offer information on user needs, technical feasibility, in addition to the competitive environment. Therefore, employing a triangulation approach involving cross – verification of insights gained from users, experts, and the already existing system provides a solid base towards the subsequent development phases. It adheres to the principle of best practice as user-centric because early involvements of stakeholders, as well as detailed studies of present landscape are crucial for successful product development.

### 3.3: Software Development Techniques

#### 3.3.1 Agile Development

Agile development is an iterative and collaborative approach that emphasizes flexibility and adaptability in software development. It breaks the development process into small, manageable increments called sprints, each typically lasting two to four weeks. This technique encourages regular feedback from stakeholders, allowing for continuous refinement of features and requirements throughout the development cycle. [33]

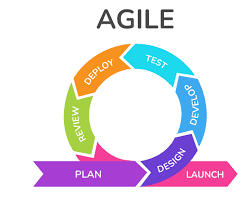


figure 1:agile methodology illustration [34]

**Advantages**

* Flexibility and Adaptability: Agile development accommodates changes in requirements, fostering a responsive and flexible development process that can readily adapt to evolving project needs.
* Stakeholder Collaboration: Regular interactions with stakeholders ensure that the project aligns with user expectations, reducing the likelihood of misunderstandings and enhancing overall user satisfaction.
* Early Delivery of Value: The incremental nature of Agile development allows for the early delivery of functional components, providing tangible value to users and stakeholders even in the early stages of development. [35]

**Disadvantages***:*

* Dependency on User Availability: Continuous stakeholder involvement may be challenging if users are not consistently available for feedback, potentially slowing down the development process.
* Documentation Challenges: The emphasis on working software sometimes leads to minimal documentation, which could pose challenges for future maintenance or for onboarding new team members.
* Scope Creep: The flexibility of Agile development may lead to scope creep if not carefully managed, potentially impacting project timelines and goals. [36]

#### 3.3.2 Waterfall Model:

The Waterfall model is a linear and sequential approach to software development, where each phase must be completed before moving on to the next. It consists of distinct stages, including requirements, design, implementation, testing, deployment, and maintenance. This technique is characterized by its structured and systematic progression through these phases.

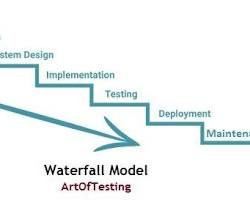


Figure 2:Waterfall Model [37]

**Advantages**:

* Clear Project Milestones: The structured nature of the Waterfall model provides clear milestones for each phase, facilitating better project management and progress tracking.
* Comprehensive Documentation: Each phase is typically accompanied by detailed documentation, ensuring a comprehensive record of project specifications, design decisions, and code implementations.
* Ease of Management: The linear progression simplifies project management, making it easier to plan and allocate resources for each phase.

**Disadvantages***:*

* Limited Flexibility: The rigid sequence of phases may lead to challenges when there is a need for changes or adaptations after the project has commenced.
* Delayed User Feedback: Stakeholder feedback is typically sought after the entire system is developed, potentially resulting in delayed identification of issues or misalignments with user expectations.
* Long Time-to-Market: The sequential nature of the Waterfall model can lead to longer development cycles, delaying the time it takes to bring a product to market. [38]

#### 3.3.3 Prototyping

Prototyping involves the creation of a preliminary version of the software to gather feedback and refine requirements before the full-scale development process begins. It allows stakeholders to interact with a tangible representation of the system, providing insights into its functionality and design.

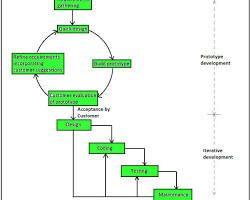


Figure 3:Prototyping Illustration [39]

**Advantages:**

* Early User Feedback: Prototyping enables the collection of early feedback, allowing users to interact with a working model and provide insights that can inform subsequent development phases.
* Clarity in Requirements: The iterative nature of prototyping helps in refining and clarifying requirements, reducing the likelihood of misunderstandings between developers and stakeholders.
* Risk Reduction: Identifying and addressing potential issues in the early stages through prototyping can significantly reduce risks associated with misunderstanding requirements or design preferences.

**Disadvantages:**

* Time-Consuming: The creation of prototypes can be time-consuming, potentially impacting project timelines, especially if significant changes are needed after user feedback.
* Potential for Scope Creep: Frequent changes during prototyping may lead to scope creep, necessitating careful management to ensure the project stays within its defined scope.
* Not Suitable for All Projects: Prototyping is most effective for projects with a degree of complexity that justifies the additional time and resources required for iterative development; it may not be suitable for straightforward projects with well-defined requirements.[40]

#### 3.3.4 Justification

The selection of software development techniques is a pivotal decision in shaping the trajectory of the project. The Agile development approach aligns with the iterative nature of the proposed web browser extension, allowing for flexibility and continuous refinement based on user feedback. This methodology is particularly suited for a project where the requirements may evolve, and regular stakeholder collaboration is crucial for ensuring the extension meets user expectations.

The consideration of the Waterfall model is beneficial for its structured and milestone-driven approach. While its linear progression may be less adaptable to changing requirements, the methodical documentation at each stage aligns well with the need for clear project management and a comprehensive record of design decisions.

Prototyping is justified in the context of this project for its potential to mitigate risks and enhance clarity in requirements. By allowing stakeholders to interact with a tangible representation of the extension early in the development process, prototyping aligns with the goal of refining features and design preferences before full-scale development.

The justification for these software development techniques lies in their compatibility with the project's objectives, each offering distinct advantages that cater to different aspects of the development lifecycle. The synthesis of Agile's flexibility, Waterfall's structured approach, and Prototyping's risk mitigation collectively contributes to a robust and adaptive software development strategy for the successful realization of the proposed web browser extension.

## 

### 3.4: System Requirements

#### 3.4.1 Software Requirements

The software requirements for the development of the AI-powered web browser extension include:

* Web Browser SDKs: To facilitate the integration of the extension into major web browsers, compatibility with Chrome, Firefox, Safari, and Edge SDKs is essential.
* JavaScript and Python Programming Environments: Utilized for both frontend and backend development, JavaScript is required for browser extension development, while Python will be employed for AI model training and implementation.
* TensorFlow.js and OpenCV Libraries: These libraries are crucial for implementing the AI image recognition features within the extension.
* WebExtension API: To interact with browser functionalities and inject scripts into web pages, the WebExtension API for each supported browser is necessary.
* Version Control System (e.g., Git): Employed for collaborative development, version control is essential to track changes, manage codebase, and facilitate team collaboration.

#### 3.4.2 Hardware Requirements

The hardware requirements for developing and testing the web browser extension are:

* Standard Development Machine: A computer with a modern processor, sufficient RAM, and storage capacity for software development activities.
* Devices for Compatibility Testing: Ensure compatibility with a range of devices, including both high-end computers and less powerful devices such as laptops and tablets.

#### 3.4.3 Functional Requirements

Functional requirements define the core capabilities of the web browser extension: [41]

* Automated Image Description: The extension should employ AI algorithms to automatically analyze and describe images on web pages.
* User Customization: Users should be able to customize the level of detail in image descriptions based on their preferences.
* Real-Time Processing: Image analysis and description should occur in real-time as users navigate web pages.
* Offline Mode: The extension should have an offline mode, enabling it to provide image descriptions based on locally stored data when internet connectivity is limited.
* Continuous Learning: Mechanisms for continuous learning and improvement should be implemented, incorporating user feedback to enhance the accuracy of image descriptions.

#### 3.4.4 Non-Functional Requirements

Non-functional requirements define aspects that are not directly related to specific behaviors but are critical for the overall performance and user experience: [42]

* Performance: The extension should operate efficiently, providing fast and responsive image descriptions without significant impact on browser performance.
* Security: Robust measures should be in place to ensure user data privacy and protect against potential security threats.
* Compatibility: The extension should be compatible with major web browsers, ensuring a seamless experience for users regardless of their browser preference.
* Usability: The user interface should be intuitive, providing a positive and accessible experience for users with visual impairments and sighted users alike.
* Maintainability: The codebase should be well-documented and modular, facilitating future updates, maintenance, and potential collaboration with other developers.
* Scalability: The extension should be designed to handle potential increases in user base and data volume without a significant degradation in performance.

### 3.5 Conclusion

In conclusion, this chapter has provided a detailed outline of the research methodology and system requirements for the development of the AI-powered web browser extension. The selected data collection techniques, including user surveys, expert interviews, and comparative analysis, lay the groundwork for an informed and user-centric design. The software development techniques, namely Agile development, the Waterfall model, and prototyping, have been justified based on their compatibility with the project's objectives, offering a balanced approach to development. The delineation of system requirements, encompassing software, hardware, functional, and non-functional aspects, provides a comprehensive guide for the subsequent stages of design, development, and evaluation. These elements collectively form the framework for the successful execution of the proposed project, ensuring its alignment with user needs, technical feasibility, and industry best practices.

**Chapter 4: System design, Implementation and Testing.**

**4.1 Introduction**

System design is the process of defining the technical roadmap for an AI-powered image captioning extension. This chapter dives into the architecture, components, interfaces, data flow, and how these aspects will be implemented and tested to ensure a robust and user-friendly extension.

**4.2 System design.**

The image captioning extension will adopt a client-server architecture to achieve its functionality. This design separates the user interface (frontend) from the image processing logic (backend).

This process is iterative, allowing for ongoing refinement as needed. The core functionalities of the frontend and backend will be defined, and the communication flow between them will be established. The following sections will detail the design of each component.

4.2.1 Logical Design

The logical design of the image captioning extension focuses on the high-level data flows and processes involved, without delving into the specific implementation details. This is often represented through a conceptual model, like the one depicted in the figure.

The logical design clarifies the following aspects:

Inputs: The system accepts an image URL from the user as input.

Data Processing: The frontend transmits the image URL to the backend server.

Backend Processes:

The backend server retrieves the image data from the provided URL.

The image data is pre-processed (e.g., resizing, format conversion) for compatibility with the image captioning model.

The pre-processed image data is fed into the image captioning model.

The model generates a caption for the image.

Outputs: The generated caption is sent back from the backend to the frontend.

Frontend: The frontend displays the caption to the user.

This logical design serves as a blueprint for the system's implementation, ensuring a clear separation of concerns between the frontend and backend components.

# APPENDIX

## References:

[1] Introduction to Web Accessibility by W3C Web Accessibility Initiative (WAI)

[2] Press release announcing the development of an AI-powered web browser extension (online article):  
[3] Create product descriptions and marketing materials":\*\* This information comes from the website of Gizzmo: <https://gizzmo.ai/>

[4] American Foundation for the Blind. (2023). Technology and visual impairment. <https://www.afb.org/>

[5] Web Content Accessibility Guidelines (WCAG) 2.2: <https://www.w3.org/WAI/>

[6] Kurniawan, S., & Wilkinson, K. (2019). The impact of alt text on the web experience for people with visual impairments. Journal of Web Accessibility, 10(1), 4.

[7] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[8] Floridi, Luciano. The ethics of artificial intelligence. Oxford University Press, 2013. https://mit-serc.pubpub.org/pub/potential-sources-of-harm-throughout-the-machine-learning-life-cycle

[9] Amodei, Dario, et al. "Concrete problems in AI safety." arXiv preprint arXiv:1606.06565 (2016).

[10] (Karpathy & Fei-Fei, 2015; Rajpurkar et al., 2020).

[11] Karpathy, A., & Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3128-3137).

[12] Google AI. (2023). Image Captioning by Google AI. Retrieved from <https://support.google.com/chrome/answer/10538231?hl=en>

[13] Proceedings of the Association for Information Science and Technology, 47(1), 1-10.

.[14] Karpathy, A., & Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3128-3137).

[15] Microsoft Cognitive Services. (2023). Image Captioning. Retrieved from <https://westus.dev.cognitive.microsoft.com/docs/services/TextAnalytics.V2.0/>

[16] Microsoft Azure. (2023, October 10). Azure Cognitive Services has achieved human parity in image captioning. <https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/concept-describe-images-40>

[17]https://www.researchgate.net/publication/338026312\_A\_deep\_learning\_approach\_for\_speaker\_recognition

[18] Clarifai (2023). Image Captioning. Retrieved from <https://www.clarifai.com/blog/clarifai-release-8.4>

[19] Clarifai. (2023, November 1). Clarifai Image Captioning: Product Documentation. <https://clarifai.com/clarifai/main/models/general-english-image-caption-clip>

[20] Existing Software Design and Development Tools

[21] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016, March). TensorFlow: A system for large-scale machine learning.

[26] Eaton, S., Kort, P., & Elston, J. (2010). The impact of respondent uncertainty on survey responses: A test of the uncertainty-response consistency hypothesis.

[27] Pointerpro. (2023, June 21). 12 Advantages and disadvantages of questionnaires. https://pointerpro.com/blog/questionnaire-pros-and-cons/

[28] Beckman, S. E., & Potter, B. J. (2015). *The art and science of expert interviewing*. Sage Publications.

[29] Kvale, S. (1996). InterViews: An Introduction to Qualitative Research Interviewing.

[30] Kvale, S. (1996). *InterViews: An introduction to qualitative research interviewing*. Sage Publications.

[31] *Comparative analysis of existing MES systems.* (2018). Journal of Manufacturing Technology Management, 29(6), 1229-1249.

[32] *Comparative analysis with existing systems.* (2018). In 2018 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW). IEEE.

[33] Beck, K. (2002). Agile software development: Principles, patterns, and practices. Addison-Wesley.

[34] <https://wisatasurabaya.id/tahapan-keuntungan-serta-prinsip-metode-agile/>

[35] Beck, K. (2002). Agile software development: Principles, patterns, and practices. Addison-Wesley.

[36] (Karpathy & Fei-Fei, 2015; Rajpurkar et al., 2020).

[37] Karpathy, A., & Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3128-3137).

[38] Royce, W. W. (1970). Managing the development of large software systems: Why waterfalls are dangerous. In Proceedings of the second ACM symposium on the implementation of programming languages (pp. 328-333).

[39] <https://www.geeksforgeeks.org/software-engineering-phases-prototyping-model-set-2/>

[40] Nielsen, J. (1993). Usability Engineering: Interaction design and usability testing. Morgan

[41] International Organization for Standardization. (2011). ISO/IEC/IEEE 24765:2010 Systems and software engineering — Vocabulary. Geneva: ISO.

[42] Wiegers, K. E., & Beatty, P. C. (2013). Software requirements. Microsoft Press.

## Budget

|  |  |  |
| --- | --- | --- |
| **Phase** | **Expense Category** | **Estimated Cost** |
| **Research and Development (2 months)** | Cloud Computing Resources | Ksh 150/month |
|  | Development Tools and Libraries | Ksh 3000 (one-time) |
|  | Literature Review and Research Materials | Ksh 500 |
| **Total** |  | Ksh 3950 |
| **Development and Implementation (3 months)** | Developer Time | No cost |
|  | User Interface Design | No cost |
|  | Web Development Frameworks | Ksh 1500 (one-time) |
|  | Accessibility Testing Tools | Ksh 150 (one-time) |
| **Total** |  | Ksh 1650 |
| **Testing and Deployment (1 month)** | Beta Testing Incentives | Ksh 100 (optional) |
|  | Bug Tracking and Resolution | Ksh 100 (estimated) |
| **Total** |  | Ksh 200 |
| **Documentation** | Printing documentation | Ksh 1000 |
| **Total** |  | Ksh 1000 |
| **Grand Total** |  | **Ksh 6800** |

Table 1: Budget

## Activity Schedule

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ACTIVITY | SEPTEMBER | | OCT | NOV | JAN | FEB | MARCH |
| Project identification |  |  |  |  |  |  |  |
| System analysis |  |  |  |  |  |  |  |
| System Design |  | |  |  |  |  |  |
| Coding and Testing |  | |  |  |  |  |  |
| Implementation |  | |  |  |  |  |  |
| Documentation |  | |  |  |  |  |  |

Table 2:Schedule/ Timeline