AI for Alt Text at WinVinaya InfoSystems

Hari P*

27 January, 2025

*Representing WinVinaya InfoSystems.

Contents

1 Abstract

This paper presents a comprehensive approach to automating the extraction, analysis, and generation of alternative text (alt-text) for images embedded in Docx documents. By leveraging machine learning models such as Llama 3.2-11B Vision, the study aims to improve document accessibility and streamline data management processes. We propose a methodology where images are first extracted from the document and assessed for the presence of alt-text. The results are meticulously recorded in an Excel file. Subsequently, Llama 3.2-11B Vision generates appropriate alt-text for images lacking descriptions, ensuring accurate context delivery and ease of comprehension for all users. This innovation not only augments document remediation efforts but also exemplifies the integration of artificial intelligence to enhance digital inclusion.

2 Introduction

Document accessibility is a critical component of digital inclusion, particularly for individuals with visual impairments who rely on assistive technologies. Alternative text (alt-text) plays a vital role in ensuring images within documents convey meaningful context. Despite its importance, the manual creation of alt-text is labor-intensive and prone to inconsistencies. This research explores the use of AI-driven methodologies to automate alt-text generation, specifically employing Llama 3.2-11B Vision, a state-of-the-art machine learning model designed for visual understanding and language generation.

3 Methodology

3.1 Overview

The methodology involves five key phases: image extraction, alt-text identification, status recording, alt-text generation, and document reintegration. Each phase employs advanced libraries and tools, coupled with Llama 3.2-11B Vision's capabilities, to deliver a seamless and efficient workflow.

3.2 Technological Framework

- **Programming Language:** Python 3.10
- **Libraries:**
  - `python-docx` for image extraction
  - `openpyxl` for Excel file management
  - `GroqClient` for integrating the Llama 3.2-11B Vision model via Groq API
- **Hardware:** High-performance GPU-enabled systems for model inference

4 Execution Plan

Step 1: Extract Images from the Document

Using the `python-docx` library, images embedded within the Docx document are extracted and stored locally. Each image is assigned a unique identifier for easy tracking.

Step 2: Assess Presence of Alt-Text

Each image is evaluated for existing alt-text using metadata inspection. If alt-text is present, it is recorded; otherwise, the image is marked for further processing.

Step 3: Record Status in an Excel File

The findings from Step 2 are logged in an Excel file using `openpyxl`. The file contains the following columns:

- **Image ID:** Unique identifier for each image
- **Alt-Text Status:** Indicates whether alt-text is present or absent
- **Existing Alt-Text:** If applicable, the alt-text is recorded

Step 4: Generate Alt-Text Using Llama 3.2-11B Vision

Images lacking alt-text are passed through the Llama 3.2-11B Vision model via Groq API. The model generates accurate and context-aware alt-text descriptions, which are reviewed for quality and relevance.

Step 5: Integrate Alt-Text Back into the Document

The newly generated alt-text is integrated back into the original document at their respective image positions using the `python-docx` library.

5 Results

5.1 Alt-Text Coverage

Initial testing on a sample dataset showed a significant improvement in alt-text coverage, with high accuracy for images requiring AI-generated descriptions.

5.2 Time Efficiency

The automated workflow reduced the time required for document remediation compared to manual alt-text generation.

5.3 Quality Assessment

Human evaluators rated the AI-generated alt-text highly for accuracy, relevance, and contextual understanding.

6 Discussion

The findings validate the hypothesis that integrating machine learning models like Llama 3.2-11B Vision can automate alt-text generation effectively. The proposed methodology demonstrated scalability, adaptability, and significant time savings. Challenges such as handling ambiguous or highly abstract images require further refinement in model training.

7 Conclusion

The integration of AI for alt-text generation significantly enhances document accessibility and reduces manual effort in remediation processes. Llama 3.2-11B Vision's capabilities illustrate the potential of AI in bridging accessibility gaps and promoting digital inclusivity.

8 Future Work

Future research will focus on:

- Expanding the model's training dataset to handle niche and industry-specific image contexts
- Integrating additional languages for multilingual support
- Developing a user-friendly interface for non-technical users

9 References

- Meta. "Llama 3.2-11B Vision Model Overview." *Meta Documentation*, 2025.
- Python Software Foundation. *python-docx Library Documentation.*
- Python Software Foundation. *openpyxl Library Documentation.*