free ebook foundation

Using AI for Accessible Alt-Text

Our Solution

Part 1: Alt-Text PyPi Package

A Python package capable of...
- Fetching and assigning data from images in HTML documents.
- Generating alt-text for images.

Super easy-to-use...
- Documentation for all functions
- **Generate alt-text for all images in one document in just five (5) lines of code!**

pip install alt-text

Out-of-the-box support for popular technologies and services...

Replicate
(Giving access to Blip, MiniGPT4, and more!)

OpenAI
(Giving access to GPT-3.5 Turbo, GPT-4.0, etc.)

Google's VertexAI *(Image captioning)*

Tesseract OCR *(Optical Character Recognition)*

PrivateGPT *(Llama2)*

Part 2: Alt-Text API

A Web API capable of...
- The same functionality as the PyPi package.
- Integrating with any web capable technology.
- Storing intermediate and final generated alt-text results.

Super easy-to-use...
- Uses REST format
- Documented using OpenAPI standards
- **Generates alt-text for all images in a document in just two (2) requests!**

The Problem

Alternative text (aka alt-text) is used to make web pages and ebooks more accessible.

Alt-text describes what an image shows for users who cannot see the image.

However, alt-text isn't always implemented when images are present.

The Client - Project Gutenberg

A volunteer effort dedicated to digitizing and distributing public domain works.

They have a database of over 70,000 ebooks for anyone to access.

The issue is, roughly 470,000* images in those ebooks lack alt-text making them inaccessible to visually impaired users.

**Not including images with bad alt-text entries*

Images in HTML
`<img src="url" alt="alternatetext">`

✗ Bad Alt-Text

alt=""

alt="[image]"

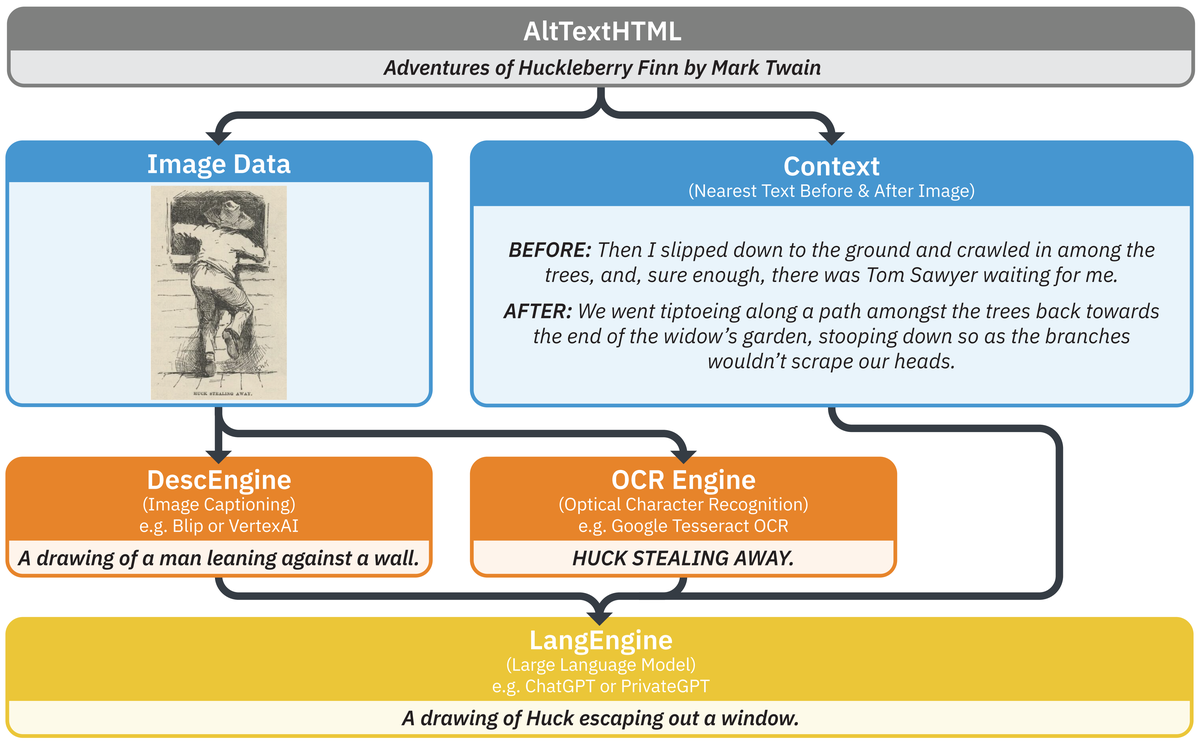
alt="no description"

✓ Good Alt-Text

alt="A black and white photo of a young man in working clothes standing in front a wall of logs."

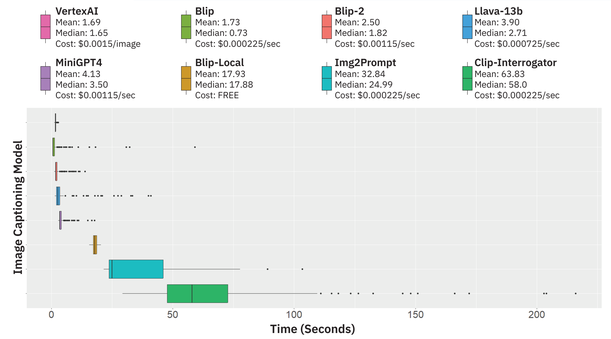
The Goal

Our goal is to develop a software solution using current AI technologies, **such as machine vision and large language models**, to assist in creating alt-text for ebooks.

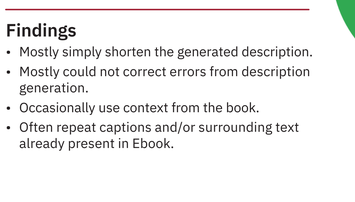


Results

Description Generation



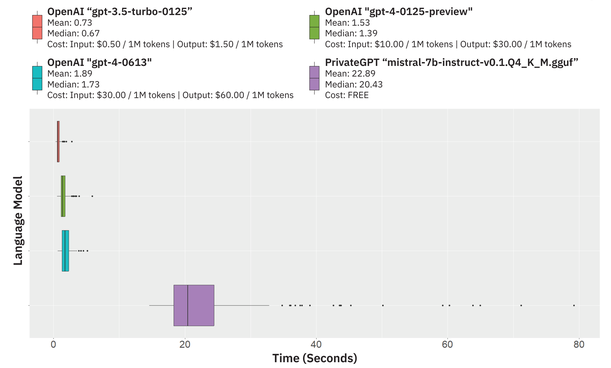
Generated Description Preference



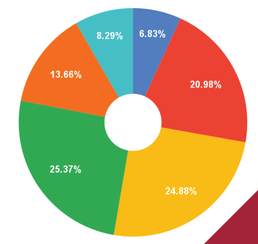
Findings
- Highly focus on text when found.
- Tend to over describe images that serve to be functional (e.g. drop caps).
- Fail to recognize when an image has more than one scene (e.g. 2 pictures in 1 image).
- Horrible at accurately describing diagrams.
- Work best with photographs over any other type of image.

Description Refinement

Alt-text description refinement using generated descriptions from the Llava-13b model. The refined texts are compared to the pre-existing alt-text, if there is any.



Refined Description Preference



Findings
- Mostly simply shorten the generated description.
- Mostly could not correct errors from description generation.
- Occasionally use context from the book.
- Often repeat captions and/or surrounding text already present in Ebook.

Jack Byrne
David Cruz
Jared Donnelly
Ethan Kleschinsky
Tyler Lane
Carson Lee