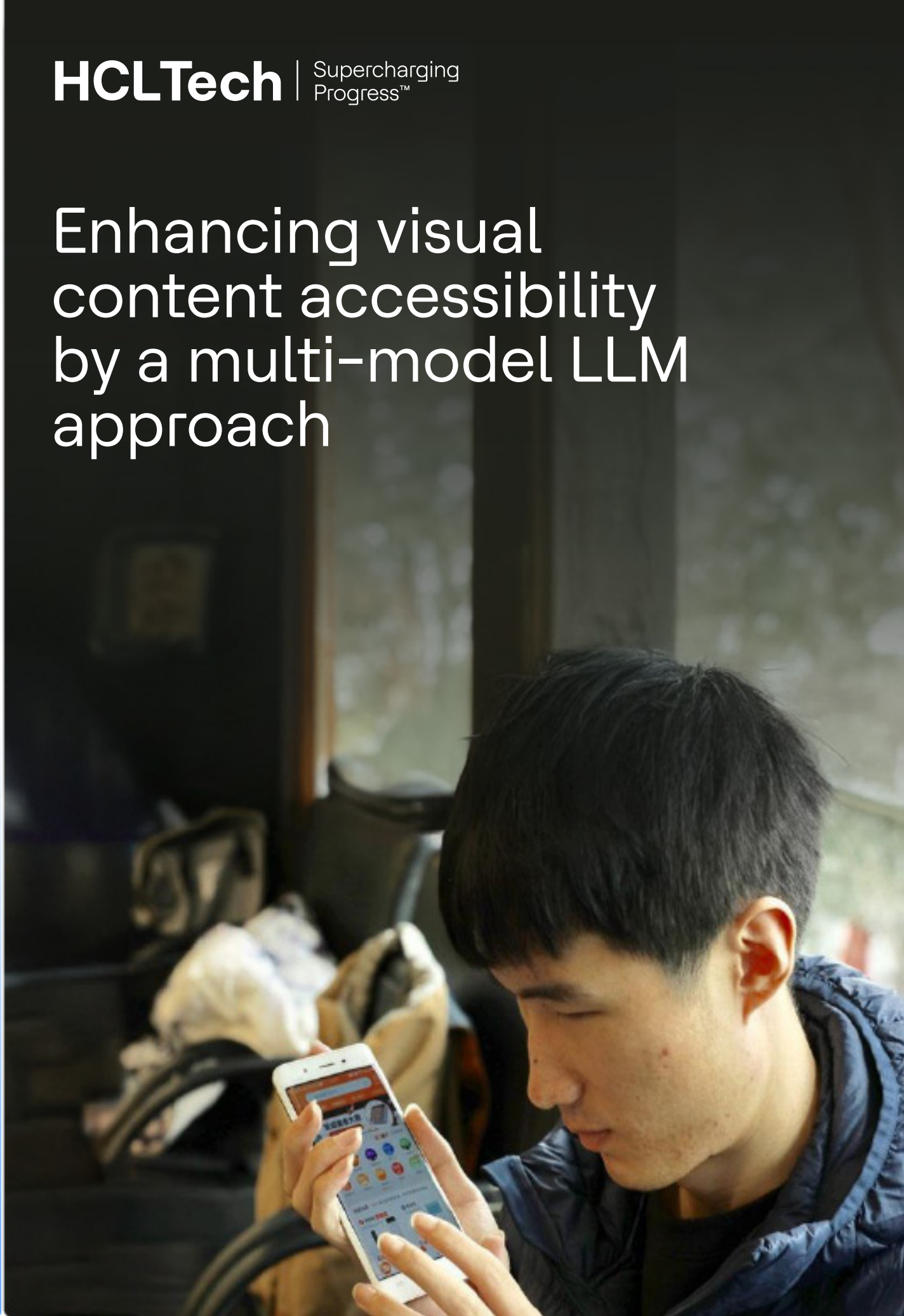


# Enhancing visual content accessibility by a multi-model LLM approach



# Table of content

<b>Abbreviations</b>	<b>3</b>
<b>Introduction</b>	<b>3</b>
<b>Business challenges</b>	<b>3</b>
<b>Problem statement</b>	<b>4</b>
<b>Related approaches</b>	<b>5</b>
<b>Solution</b>	<b>6</b>
<b>Benefits</b>	<b>8</b>
<b>Discussions</b>	<b>9</b>
<b>Conclusion</b>	<b>10</b>
<b>References</b>	<b>10</b>
<b>Author information</b>	<b>11</b>

# Abbreviations

BLEU	Bilingual Evaluation Understudy
SEO	Search Engine Optimization
BP	Brevity Penalty
METEOR	Metric for Evaluation of Translation with Explicit Ordering
CIDE	Consensus-based Image Description Evaluation

## Introduction

Web accessibility is paramount for ensuring digital content is usable by individuals with disabilities. A critical component of accessible web design is providing alternative text (alt text) for images. Alt text serves as a textual description of an image, enabling screen reader users to understand the visual content. Accurate and Descriptive alt text helps users comprehend the content and context of web pages.

## Business challenges

### The need for efficient and accurate alt text generation

The increasing volume of visual content on the web has created a significant challenge for content creators to cope with the growing emphasis on digital accessibility. Generating accurate and descriptive alt text for every image manually is time-consuming and resource-intensive.

### Key business challenges include:

- **Time and cost:** The manual process of creating alt text is labor-intensive and can impact content creation timelines and budgets.
- **Inconsistent quality:** Manual alt text creation may lead to inconsistencies in quality and accuracy, affecting user experience and accessibility.
- **Legal and compliance risks:** Failure to provide accurate alt text can result in legal and reputational risks.
- **Search Engine Optimization (SEO):** Descriptive alt text is crucial for image search engine optimization, impacting website visibility and traffic.

By automating the alt text generation process, businesses can address these challenges, improve website accessibility and enhance overall user experience.

# Problem statement

## The challenge of accurate alt-text generation

Ensuring web accessibility for visually impaired users is a critical challenge. Providing accurate and descriptive alt text for images is essential for conveying visual information to screen reader users.

Existing methods, such as rule-based systems and traditional machine learning (ML) approaches, have limitations in accurately capturing image content and generating contextually relevant alt text. This research aims to address these challenges by leveraging the capabilities of large language models (LLMs) for automated alt-text generation.

## The promise of automation

This paper presents a novel approach to automating the generation of alt text using LLMs to address the challenges associated with alt text creation. By leveraging the capabilities of advanced language models, we aim to improve the efficiency and accuracy of alt text generation, ultimately improving web accessibility and enhancing the user experience for users with visual impairments.

The increasing volume of visual content on the web has created a significant challenge for content creators to cope with the growing emphasis on digital accessibility. Generating accurate and descriptive alt text for every image manually is time-consuming and resource-intensive.

## Sample examples:



Source: [Accessibility - Wikipedia](#)

**Alt Text:** Not provided

**Problem:** The user is not able to understand the purpose of the image

**Recommended Alt Text:** A group of people are sitting, wearing headphones, with musical instruments in the room



**Source:** [Accessibility - Wikipedia](#)

**Alt Text:** Business owner with a construction helmet

**Problem:** The given Alt Text does not match with the context

**Recommended Alt Text:** A woman is helping another woman with a craft project

## Related approaches

The significance of accurate alt text for visually impaired users has led to the development of various approaches to automate the generation process. While progress has been made, challenges persist in creating alt text that accurately reflects image content and context.

### Manual alt text creation

Traditionally, content creators or web developers have generated alt text manually. While this approach ensures accuracy when done diligently, it is time-consuming, especially for large websites with numerous images. Furthermore, consistency in alt text quality can vary significantly across different authors.

### Rule-based systems

Automating alt text generation using rule-based systems employed predefined rules and heuristics to generate descriptions based on image attributes such as file name, dimensions and color palette. However, the limitations of such approaches became evident as they struggled to capture the nuances of complex image content.

### ML approaches

With advancements in computer vision and natural language processing, ML techniques have been applied to alt text generation. These methods typically involve training models on large datasets of images and corresponding human-written alt text. While promising, these approaches often face challenges in handling diverse image content and generating contextually relevant descriptions. Also, training the models requires a lot of data, especially image data, which is a big challenge.

## LLMs

Recent breakthroughs in natural language processing have led to the development of powerful language models, such as Gemini. These models have demonstrated remarkable capabilities in generating human-quality text, making them suitable for tasks like alt text generation. However, their application to this specific domain is still in its early stages and further research is needed to evaluate their effectiveness.

Our proposed approach leverages the strengths of LLMs to overcome the limitations of previous methods and provide a more accurate and robust solution for alt text generation.

# Solution

## Data collection and preprocessing

A dataset of web pages was collected to develop and evaluate our alt text generation system. This dataset included diverse content to ensure the model's ability to handle various image types and contexts.

Image URLs were extracted from the HTML content of these web pages using web scraping techniques. To do this, we used the beautiful soup package. To focus on informative images, filters were applied to exclude decorative elements such as icons and others based on attributes like file format and other display properties. Hidden images were also filtered since they are not displayed in the web design.

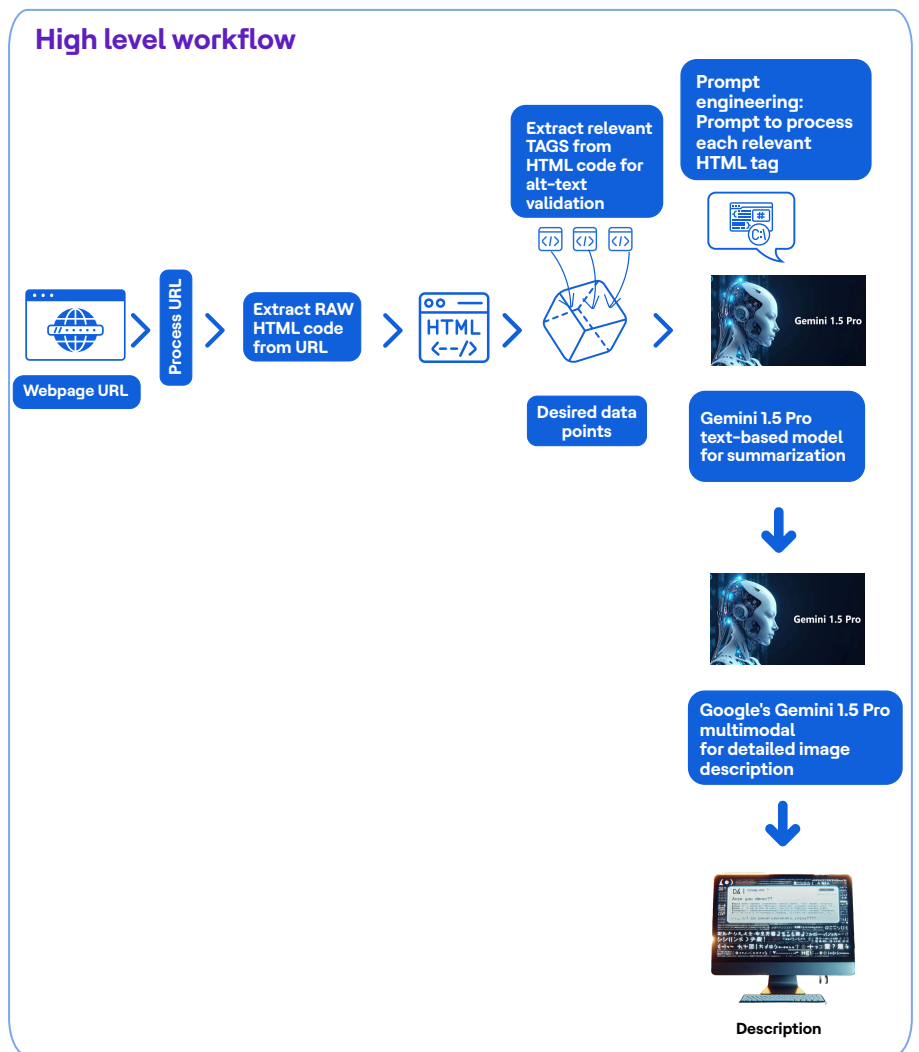
## Image extraction and processing

The extracted image URLs were used to download the corresponding images. The downloaded images were saved in a standardized format (e.g., JPEG) to ensure compatibility with the image processing pipeline. Minimal preprocessing was applied to the images, as the Gemini model is designed to handle diverse image formats and content.

## Alt text generation

The core of our approach involves leveraging the Gemini language model (Gemini 1.5 pro) to generate descriptive alt text for images. Each image was processed individually and a detailed prompt was constructed to guide the model's output. The prompt included information about the image and the relevant context derived from the surrounding HTML content.

The Gemini model was used in two different modes: the vision-based mode generated a detailed description of the image and the text-based mode summarized the detailed description obtained from the vision model.



## Evaluation methodology

To assess the quality of the generated alt text, a comprehensive evaluation framework, including quantitative and qualitative metrics, was employed.

- **Quantitative evaluation:** The BLEU score was calculated to measure the similarity between the generated alt text and human-written reference text.
- **Qualitative evaluation:** Human experts evaluated the generated alt text based on accuracy, descriptiveness and conciseness criteria.

By combining these evaluation methods, we aimed to understand the system's performance comprehensively.

## Quantitative evaluation

The primary metric used to evaluate the system was the Bilingual Evaluation Understudy (BLEU) score. This metric calculates the n-gram precision between the generated alt text and a reference human-written alt text. A higher BLEU score indicates a closer match between the two texts.

Mathematically, the BLEU score is given as follows:

$$\text{BLEU Score} = \text{BP} * \exp(\sum_{i=1}^N (w_i * \ln(p_i)))$$

Here,

**BP** stands for Brevity Penalty

**w<sub>i</sub>** is the weight for n-gram precision of order i (typically weights are equal for all i)

**p<sub>i</sub>** is the n-gram modified precision score of order i

**N** is the maximum n-gram order to consider (usually up to 4)

While the BLEU score provides a quantitative measure of similarity, it has limitations in capturing semantic nuances and context. Additional evaluation metrics such as METEOR[SS1] or CIDER[SS2] could be considered in future work to address this.

### Qualitative evaluation

Human evaluation was conducted to better understand the system's performance. A group of human evaluators assessed the quality of generated alt text based on criteria such as accuracy, descriptiveness and consistency. This qualitative analysis provided valuable insights into the system's strengths and weaknesses.

Overall, the results demonstrated the potential of the Gemini model in generating accurate and informative alt text. However, further improvements are necessary to address challenges such as handling complex images and ensuring consistency across different image domains.

## Benefits

The implementation of an automated alt text generation system offers several key benefits for website owners, content creators and users with disabilities:

- **Improved accessibility:** By providing accurate and descriptive alt text for all images, businesses can significantly enhance the accessibility of their websites for users with visual impairments.
- **Time and cost savings:** Automating the alt text generation process eliminates the need for manual effort, saving time and resources.
- **Consistency:** Automated systems can ensure consistent alt text quality across the entire website, reducing the risk of errors or omissions.
- **Enhanced user experience:** Accurate alt text improves the overall user experience by providing context-based descriptions for the images.
- **Legal compliance:** Adhering to accessibility standards and regulations can help businesses avoid legal issues and reputational damage.

By addressing these challenges and opportunities, automated alt-text generation can contribute to a more inclusive and accessible digital landscape.

# Discussions

Integrating LLMs for alt text generation presents a promising approach to enhancing web accessibility. Our results demonstrate the potential of the Gemini model (especially Gemini 1.5 pro) in producing accurate and informative descriptions for images. However, several challenges and opportunities for improvement have been identified.

## Limitations and challenges:

While the Gemini model has shown promising results, there are limitations to consider. Complex images with multiple subjects or abstract concepts can challenge the model, leading to less accurate or incomplete alt text. Additionally, the evaluation metrics used in this study, such as the BLEU score, provide a quantitative measure of similarity but may not fully capture the nuances of human perception.

## Future directions:

To address these limitations, future research should focus on:

- **Image preprocessing:** Exploring advanced image preprocessing techniques to extract more meaningful features for the model.
- **Model enhancement:** Fine-tuning the Gemini model on a larger and more diverse dataset to improve its ability to handle complex images.
- **Hybrid approaches:** Combining rule-based methods with the language model to address specific image types or scenarios.
- **User studies:** Conducting in-depth user studies to assess the impact of generated alt text on user experience and identify areas for improvement.

Improving these areas can enhance the alt text generation process and create a more inclusive web experience for users with visual impairments.

# Conclusion

This paper introduced an innovative method for automating the creation of alternative text for images utilizing the Gemini LLM. By leveraging the model's ability to understand and generate context-based text from images, we created a system that produces human-like alternate text for images on the web pages.

Our findings demonstrate the potential of large language models to address the challenges associated with alt text creation manually and by other means. While the system shows promise, further research is necessary to refine the model, explore additional evaluation metrics and incorporate user feedback.

By advancing the field of automated alt text generation, we contribute to a more inclusive web experience for differently abled users. Future work will develop hybrid methods, include multimodal data and address ethics for robust solutions.

# References

<https://www.w3.org/TR/WCAG20-TECHS/H37.html>

# Author information

## Vijayaraghavan. A



Vijayaraghavan is an AI/ML engineer with more than 27 years of experience in the IT industry. He holds a master's degree in computer applications and an advanced diploma in data analytics. He has strong expertise in data analytics and has successfully executed and delivered projects, particularly in developing predictive models using machine learning and deep learning technologies. Additionally, he has extensive experience with large language models (LLMs) and has completed research and development projects in generative AI.

## Manoj Kumar Nishani



Manoj is a highly regarded Accessibility Project Manager with over 17 years of experience in the technology industry. He holds a Master's degree in Computer Science along with certifications in Accessibility, such as the DHS Certified Trusted Tester and Core Competency Accessibility certifications. Manoj has extensive expertise in implementing and managing accessibility programs for both web and mobile applications.

He leads cross-functional teams to ensure compliance with WCAG 2.1, Section 508, and other accessibility standards. Known for his strategic approach and excellent communication skills, Manoj excels in advocating for inclusive design, conducting accessibility audits and fostering a culture of accessibility within organizations.

# HCLTech | Supercharging Progress™

HCLTech is a global technology company, home to 222,000+ people across 60 countries, delivering industry-leading capabilities centered around Digital, Engineering and Cloud powered by a broad portfolio of technology services and software. The company generated consolidated revenues of \$12.3 billion over the 12 months ended December 2022. To learn how we can supercharge progress for you, visit [hcltech.com](https://hcltech.com).

[hcltech.com](https://hcltech.com)

