Breaking the Linear Barrier: A Multi-Modal LLM-Based System for Navigating Complex Web Content

Gabriel Moterani, Randy Lin

IEEE International Workshop on Intelligent Tools for Accessibility – Technologies Transforming the Future of Inclusive Solutions
(ITA 2025)

July 11th, 2025

Summary

Abstract

- Visually impaired users struggle with **complex**, dynamic **websites** due to linear **screen readers** and **limited** semantic context.
- The paper proposes a **multi-modal framework** using **LLMs**, computer vision, and dynamic **DOM manipulation**.
- The system enhances **semantic clarity**, non-linear navigation, and **richer interaction** through a conversational interface.
- A **prototype** was deployed in a modern browser to test **real-world applicability**.
- **Evaluation** focused on Canada's most visited websites, including a **detailed demo** on a ticketing site.
- The **prototype** successfully helped **users understand page** content and **complete tasks** like buying concert tickets.
- The work demonstrates how combining **vision**, **language models**, and **browser** control can **improve accessibility**.
- It sets the stage for future research on performance, scalability, and personalization.

Background

Accessibility Overview

- After the creation of Internet W3C was created to build guidelines.
- **WCAG** was created in **1991.**
- WCAG is improved until version 3 (2024).

- **1.5 M users** in Canada have some kind of visual impairment.
- **17%** experience **barriers** while navigating web.
- **88%** of the **websites** are **not** considered **accessible**.

Sources: CNIB, Statican, Accessibility Checker

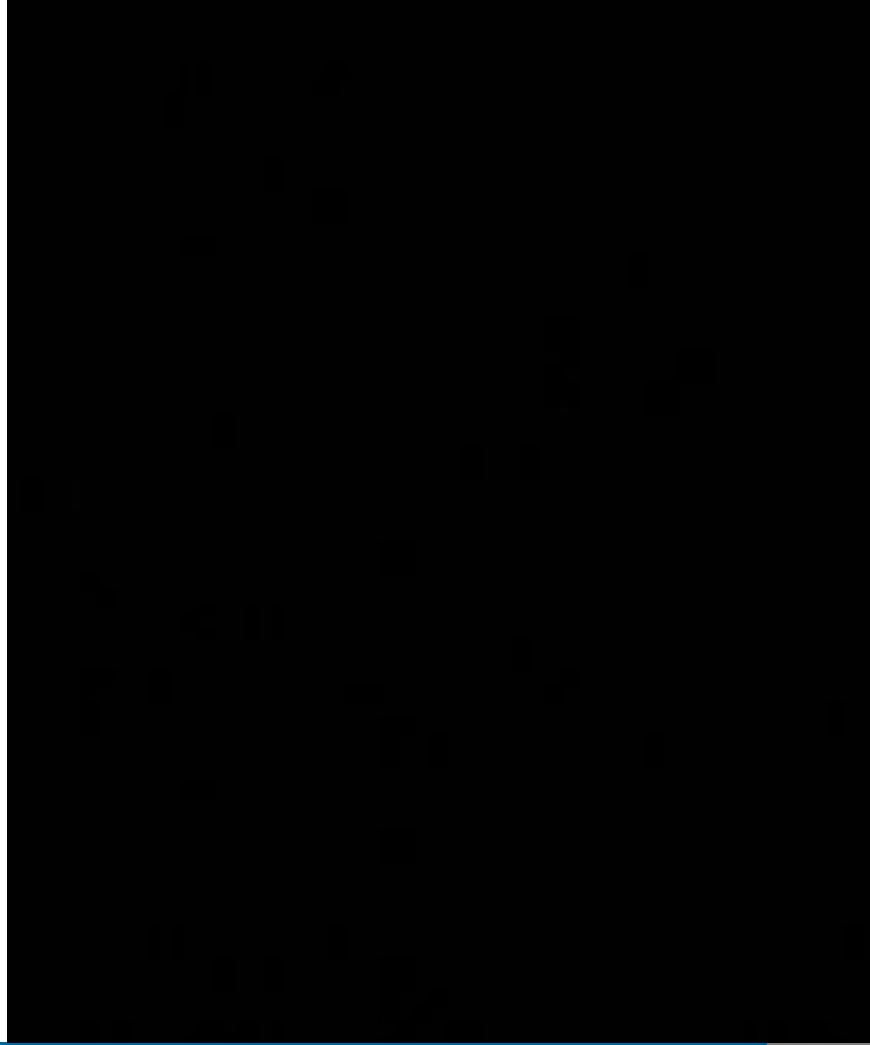
Background

Problem Demonstration

You enter a news article about an event happening in Toronto's High Park.

The first thing you hear is the alt description of the hero section image:

"A cheerful gathering of friends playing in the park, with delicious food laid out and laughter filling the air".



Background

Problem Demonstration

Missed **Context**:

- Demographic
- Type of food
- Type of activity
- Amount of people
- Others….

Background

WCAG Application

Developers don't follow guidelines:

- Lack of **knowledge.**
- Lack of **time.**
- Lack of **tools.**
- Too many **complicated** guidelines.

Sources: CNIB, Statican, Accessibility Checker

Literature Review

- **Screen readers** still force visually impaired users into **slow**, **linear** navigation, causing cognitive overload on complex pages.
- **Existing accessibility tools** (alt text, ARIA) are often **poorly developed**, missing, or misused for SEO, leaving gaps in **semantic context**.
- **AI prototypes** show promise (e.g., summarizing products, filling forms) but remain **domain-specific**, brittle, or disconnected from **normal browsing workflows**.
- **Visual compiler** approaches use **screenshots for context** but fail to integrate seamlessly with textual analysis and **typical browser use**.

1. Key **gaps** identified:
    a. Lack of **non-linear** navigation
    b. Insufficient semantic **context** for visual elements
    c. Absence of **general-purpose**, cross-site task support

Methodology

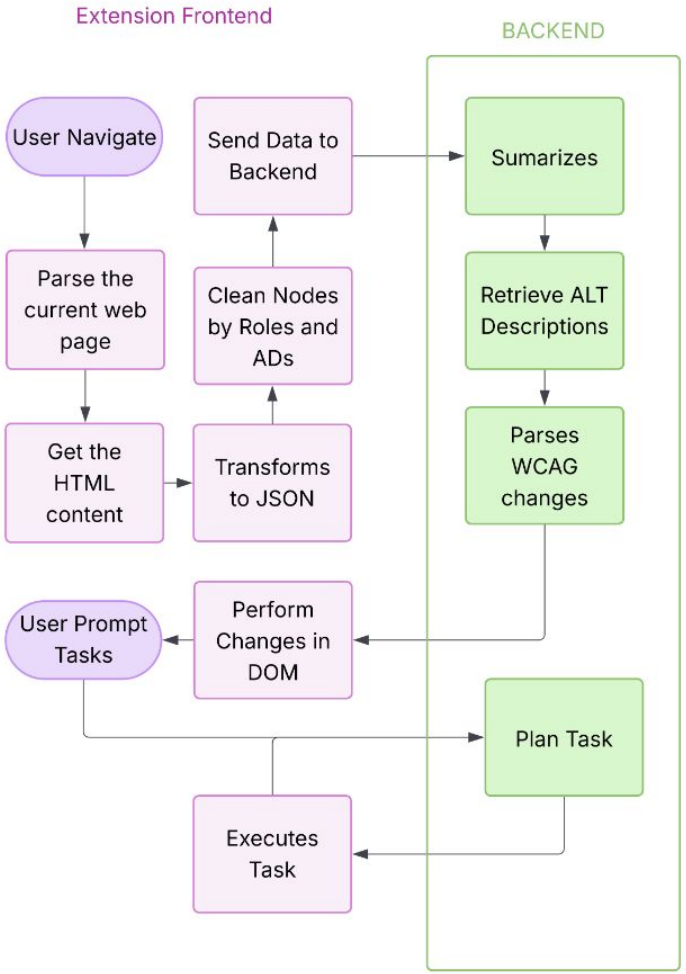
Proposed Solution

- Web browsers (e.g., Chrome, Safari) render **HTML into visual content**; Chrome leads desktop use (66%), Safari dominates mobile (22%).
- Browser extensions **enhance functionality** by interacting with the **DOM and Accessibility Tree**—used by screen readers.
- This project uses extensions to **bridge DOM and accessibility data** for richer **user interaction**.
- The system architecture includes **four cooperating modules** coordinated by a **task graph** and **state log**.
- A central **LLM handles reasoning**: it connects visual/DOM data, conversational input, and low-level user actions.

Methodology

System Overview

1. **Extension** consumes **page modifications**.
2. DOM (HTML) and Accessibility Trees are **collected**.
3. Data is **cleaned** (Improve LLM adoption).
4. Data is prepared by **types**.
5. Data is **summarized** by LLM.
6. **Images descriptions** are obtained using computer vision techniques.
7. **WCAG compliance** is analyzed.
8. DOM is **updated** w/ better **accessibility**.
9. **User** can **interact** w/ page using LLM.



Findings

1. The system was **tested** on Canada's 10 most visited sites (per SemRush) plus others, **repeating tasks** 5 times per site.
2. **Metrics collected**: HTML size, image count, WCAG violations, category, monthly traffic, and task completion efficacy.
3. Tests showed **better navigation and context**, especially on image-heavy sites; issues remained with **SVGs/videos**.
4. Moderate **correlation** found: fewer **WCAG violations** → higher **task success**; **larger HTML** size improves context and **task completion**.
5. **Model hallucinations** required careful prompt design and potential fine-tuning.
6. Used **GPT-4.1** for speed/ease, but l**imited by 200,000 tokens/minute** (tier 1); future work suggests higher tiers and **alternate models**.

Limitations

- High **computational costs** due to **real-time LLM** and **computer vision inference**, requiring powerful GPUs.
- **Added latency** from inter-module communication and **repeated page** rendering may hurt usability.
- **Limited** ability to generalize across diverse or **highly dynamic web structures**.
- **Occasional hallucinations** from the LLM leading to incomplete tasks or DOM errors.
- **Reliance on GPT-4.1**, which may lack capacity for larger datasets and deeper context.

Conclusions

- LLMs **can** help **improve** web accessibility.
- There is still a **lack** of **research** and papers focused on this **topic**.
- The **prototype** demonstrate the capability to **enhance navigation and context**.
- **Cost** and computational **latency** remain significant **barriers**.
- **Future work** will explore **multi-agent** approaches to handle diverse **structures** and **tasks**.

Thank You

Do you have any questions?

Gabriel Moterani
gamorimmoterani@algomau.ca

IEEE
Advancing Technology for Humanity