

Generative AI Facilitating Image Recognition for the Visually Impaired in Online Learning

IA generativa que facilita el reconocimiento de imágenes para personas con discapacidad visual en el aprendizaje en línea

Wei-Feng Tung^{1*}, Patrick C. K. Hung², Chen-Wei Hsieh¹, Wong Lin¹, Yen-Chen Chen¹, Meng-Chieh Lin¹

¹Fu-Jen Catholic University, New Taipei City, Taiwan

²University of Ontario Institute of Technology, Toronto, Canada
076144@mail.fju.edu.tw, patrick.hung@ontariotechu.ca

Fecha de recepción: 12 de agosto de 2024

Fecha de aceptación: 9 de diciembre de 2025

Summary. Visually impaired learners often face significant challenges when accessing online educational videos, particularly programming tutorials that rely heavily on visual elements. Traditional teaching resources rarely consider accessible formats, resulting in limited participation of visually impaired students in STEM learning. This study proposes an Artificial Intelligence (AI)-powered system called Image Narrator, which incorporates a Chrome extension, YouTube plugin, and accessible website to support visually impaired learners in understanding video-based instructional content. By leveraging advanced generative AI and computer vision technologies, the system extracts text and visual elements from video frames and converts them into accessible descriptions or screen-reader-friendly formats. A user study involving 24 participants, including visually impaired learners, was conducted at the Canadian National Institute for the Blind (CNIB). Results indicate high satisfaction with the system's image recognition, auto-generated exercises, question-and-answer functions, and overall usability, demonstrating that AI-powered visual narration can effectively reduce learning barriers and enhance independent learning for visually impaired students.

Keywords: Generative AI, Computer vision, Assistive technology, Visually impaired learners, Online learning accessibility.

Resumen. Los estudiantes con discapacidad visual suelen enfrentar desafíos significativos al acceder a videos educativos en línea, en particular a tutoriales de programación que se basan en gran medida en elementos visuales. Los recursos didácticos tradicionales rara vez consideran formatos accesibles, lo que resulta en una participación limitada de los estudiantes con discapacidad visual en el aprendizaje de STEM. Este estudio propone un sistema impulsado por Inteligencia Artificial (IA) llamado Narrador de Imágenes, que incorpora una extensión de Chrome, un complemento de YouTube y un sitio web accesible para ayudar a los estudiantes con discapacidad visual a comprender el contenido instructivo basado en video. Mediante el uso de tecnologías avanzadas de IA generativa y visión artificial, el sistema extrae texto y elementos visuales de fotogramas de video y los convierte en descripciones accesibles o formatos compatibles con lectores de pantalla. Se realizó un estudio de usuarios con 24 participantes, incluidos estudiantes con discapacidad visual, en el Instituto Nacional Canadiense para Ciegos (CNIB). Los resultados indican una alta satisfacción con el reconocimiento de imágenes, los ejercicios autogenerados, las funciones de preguntas y respuestas y la usabilidad general del sistema, lo que demuestra que la narración visual impulsada por IA puede reducir eficazmente las barreras de aprendizaje y mejorar el aprendizaje independiente de los estudiantes con discapacidad visual.

Palabras Clave: IA generativa, visión artificial, tecnología de asistencia, estudiantes con discapacidad visual, accesibilidad al aprendizaje en línea.

1 Introduction

Access to online learning materials remains a significant barrier for the visually impaired, despite advancements in artificial intelligence and digital education. Globally, over 250 million individuals are visually impaired, limiting their employment opportunities and hindering their participation in STEM education. Existing STEM curricula rely heavily on visual demonstrations, such as block-based programming and video tutorials, making it difficult for visually impaired learners to engage in mainstream instruction.

Tools such as tactile systems and Microsoft's Code Jumper have attempted to address these challenges; however, most programming education tools still lack complete accessibility support. Studies consistently indicate that visually impaired learners face barriers related to navigation, formatting, and access to digital content, often relying on interpersonal assistance for information acquisition.

Recent advancements in AI—including machine learning, deep learning, and computer vision—enable new possibilities for assistive learning tools. These include image recognition, audio narration, and content interpretation. While companies such as Microsoft, NavCog, and Horus have developed AI-based mobility or object-recognition tools, few focus on improving access to educational resources or decoding instructional videos.

Therefore, there is a critical need for specialized AI tools that support the academic learning needs of visually impaired students.

This study presents *Image Narrator*, an AI-enabled learning support system designed to extract, interpret, and convert visual information from online instructional videos into accessible formats. The system aims to improve the autonomy and learning effectiveness of visually impaired students in online environments.

2 Existing accessibility systems and tools

El aprovechamiento o logro escolar, se refiere al nivel de conocimientos de un alumno comparado con la norma de edad y nivel académico. Normalmente para valorarlo, éste es medido mediante una evaluación; sin embargo; como todos los problemas complejos, se trata de un fenómeno multifactorial; y aunque se manifiesta a lo largo de un periodo específico, a través de una calificación cuantitativa, no es el producto de una única aptitud.

2.1 NVDA (NonVisual Desktop Access)

NVDA is a free, open-source screen reader widely used in both educational and professional settings. It converts on-screen elements into speech or braille and provides robust support for web navigation, document reading, and software operations, offering a cost-effective accessibility solution. [5].

2.2 JAWS (Job Access With Speech)

JAWS is a commercial screen reader that provides comprehensive accessibility for tasks ranging from text reading to programming. Its extensive functionality allows visually impaired users to operate a wide range of software tools effectively. [4].

2.3 ChromeVox

ChromeVox is Google's built-in screen reader for Chrome OS and Chrome browser users. It supports keyboard navigation, speech output, structural navigation, and auditory cues (earcons), enabling blind and low-vision users to browse web content efficiently. [2].

1. Designed for Chrome OS: While it works on all platforms running Chrome, ChromeVox is primarily designed for Chromebooks.
2. Keyboard Commands: ChromeVox provides a set of keyboard shortcuts that allow users to navigate, select, and perform actions.
3. Speech Output: ChromeVox speaks the currently selected or pointed-to content and offers adjustable speech speed and volume.
4. Navigation Support: Users can quickly navigate to different parts of a webpage, such as headings, links, or lists, using specific keyboard commands.
5. Earcons: These are sound effects used to indicate different events or states, such as a new window opening or a page loading.

3 System Architecture and Demonstration

This system architecture, called 'Image Narrator', illustrates the workflow of an AI-powered educational support platform that integrates accessible web content, Chrome extension services, and an AI assistant. The architecture is organized into five conceptual layers: Presentation Layer, Flow Logic, Application Layer, Business Logic, and Data Access Layer, each representing distinct responsibilities in the system pipeline. The primary goal is to support visually impaired users or learners with special needs by recognizing video content, generating questions, and enabling intelligent interaction with AI (see Figure 1).

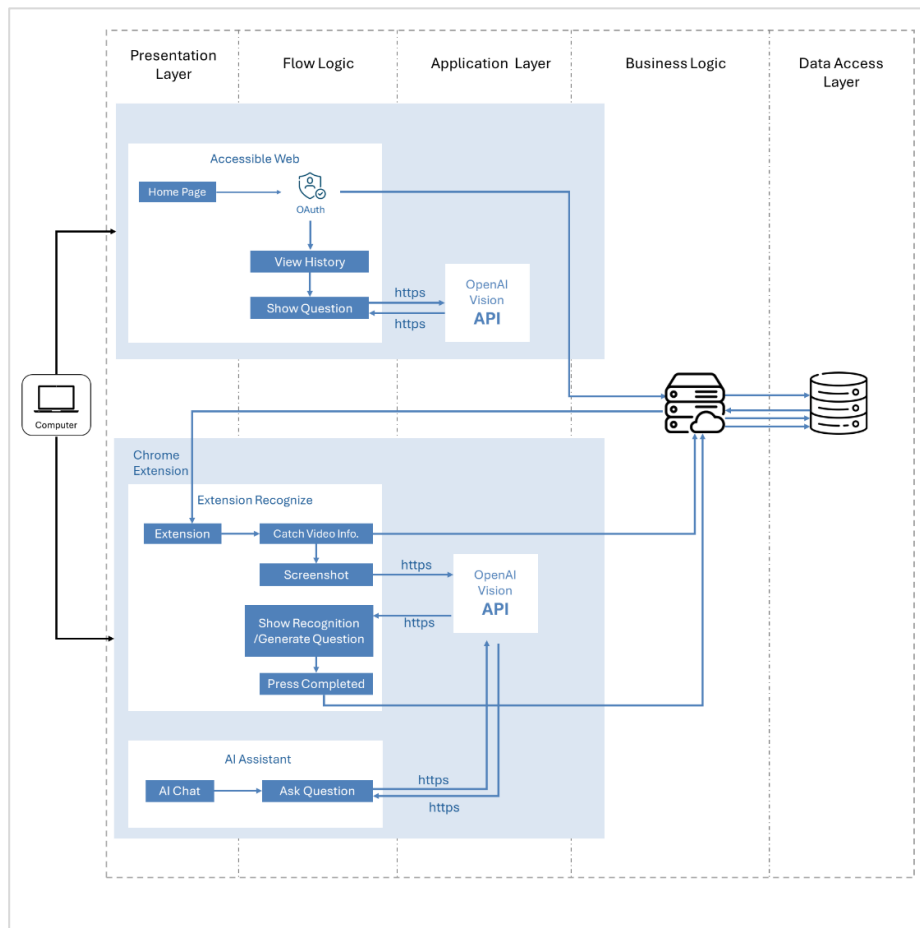


Figure 1. System Architecture of *Image Narrator*.

3.1 Presentation Layer

Includes the accessible web portal, Chrome extension, and AI assistant. Users can authenticate, view learning history, capture screenshots from YouTube videos, and ask questions through the AI assistant.

- **Accessible Web:** Users authenticate via OAuth and can access a personalized homepage where they can:
 - View previously generated questions through the "View History" function.
 - Interact with questions via the "Show Question" interface.
- **Chrome Extension:** A browser-based plugin that captures real-time video content for processing and analysis.
- **AI Assistant:** A chat-based interface allowing users to engage in natural language interactions by posing and receiving questions derived from multimedia content.

3.2 Flow Logic

Controls interactions between users and the system. The extension captures video frames, processes screenshots, sends them to the AI model, and returns recognized text and questions.

- **Accessible Web:** After user authentication, the system fetches stored question data or calls the OpenAI Vision API to interpret visual information.
- **Extension Recognize Module:**
 - Captures video content metadata via "Catch Video Info."
 - Takes screenshots at relevant intervals for analysis.
 - Sends visual content to the AI API for recognition.
 - Returns the generated question or recognized content to the user.

- Enables the user to confirm the process by pressing "Completed."
- **AI Chat:** Allows users to send a question prompt, which is processed via API calls and returns a response accordingly.

3.3 Application Layer

Handles direct communication with the OpenAI Vision API, enabling image recognition, content extraction, and question generation via HTTPS requests.

- HTTPS requests are sent from both the web interface and the Chrome extension to the Vision API.
- The API responds with recognized entities or contextual descriptions that are used to generate educational questions or assistive content automatically.

3.4 Business Logic

Manages application rules, question templates, user history retrieval, and the coordination between API results and interface elements:

- Matching visual content to appropriate question templates.
- Storing and retrieving question-answer pairs based on user history.
- Routing user interactions from various components (web, extension, assistant) to a centralized logic processor.

3.5 Data Access Layer

Stores user data, recognition outputs, question sets, and authentication information in a secure database:

- A database system for storing:
 - User credentials and OAuth tokens.
 - Visual recognition results.
 - Generated questions and user interaction history.
- Secure communication between the logic layer and the database to ensure privacy and integrity.

3.6 System Demonstration

Screenshots from instructional videos are recognized as text and narrated aloud via screen readers (Figure 2). Users can ask questions, receive AI-generated practice exercises, and maintain learning records through the accessible website.

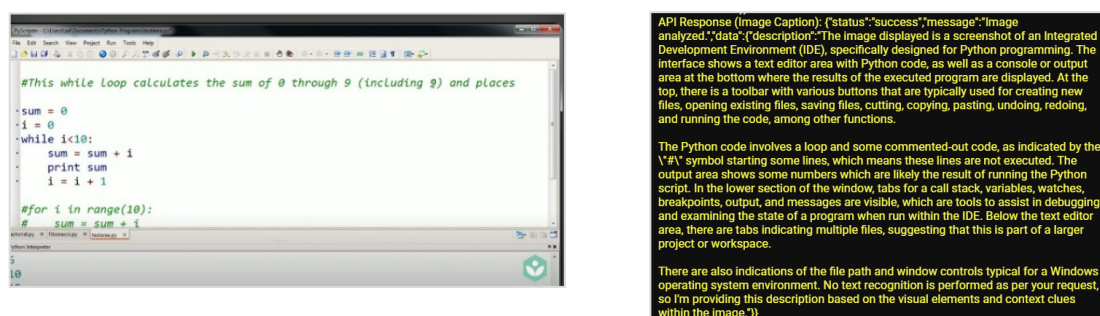


Figure 2. A screenshot of programming learning from an online video.

Users can also ask the AI assistant questions to get answers easily. Image Narrator can also provide several questions for testing based on the screenshot images the user provided before. They can practice the related questions from AI to gain a deeper understanding of the topics (see Figure 3).

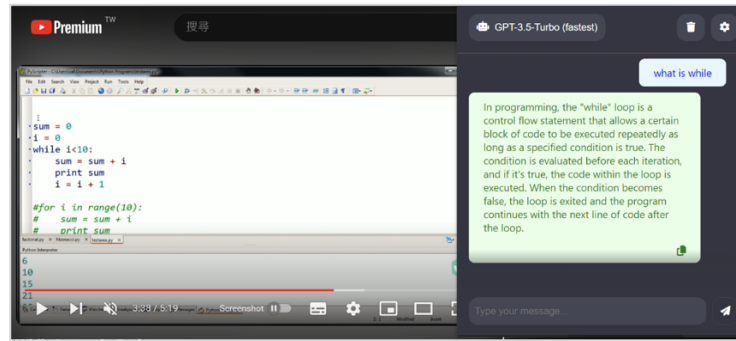


Figure 3. AI questions for practicing on Image Narrator.

This system can also help the user save learning records to help them take online notes on the website. Each screenshot will save its learning content into the database system (see Figure 4).

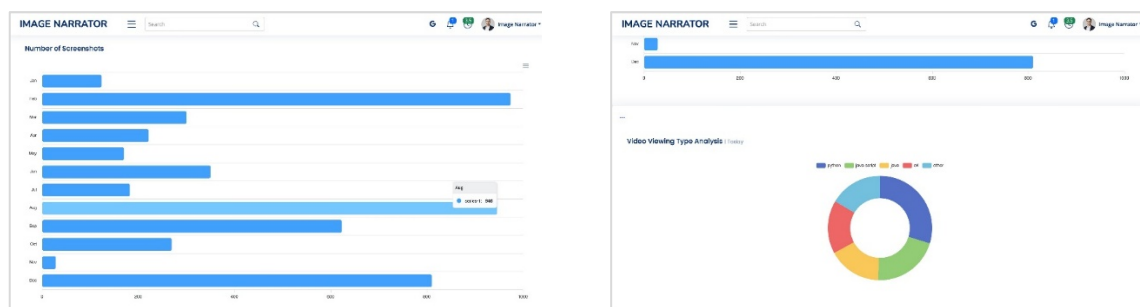


Figure 4. GenAI provides questions for practicing after learning.

4 User Testing and Results

To verify the effects of Image Narrator, some real users were tested at CNIB in Canada in 2024. 24 users have finished the experiments, including 11 males and 13 females. Among the subjects, 9 were aged 20-30 years old. 6 were 30-40 years old. 3 were 40-50 years old. 5 were 50-60 years old. 1 was aged 60-70 years old. 5 sighted people, 19 visually impaired people.

In addition to the basic information, the above users answered the questionnaire with 5 other questions.

1. How helpful is the image recognition function in assisting you in understanding the video?

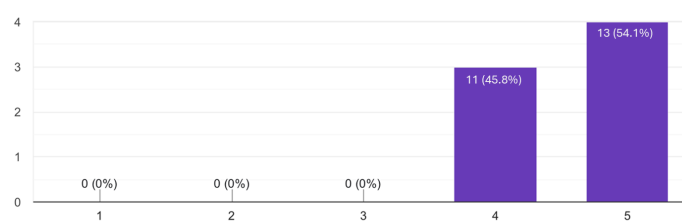


Figure 5. Results of answering Question 1 of the questionnaire after the experiment.

Rating 1-3 is 0 responses, Rating 4 is 11 responses (45.8%), and Rating 5 is 13 responses (54.1%). The mean score is 4.54, indicating a high level of agreement or satisfaction. The standard deviation is 0.50, suggesting relatively low variability in responses (see Figure 5).

2. How helpful is the exercise (generated by Image Narrator) for learning?

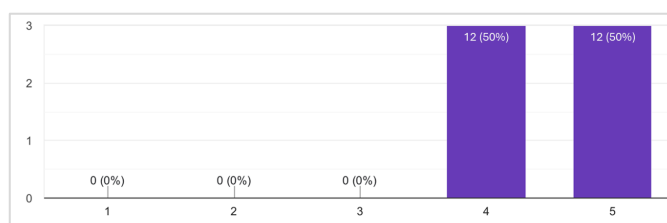


Figure 6. Results of answering Question 2 of the questionnaire after the experiment.

Rating 1-3 is 0 responses, Rating 4 is 12 responses (50%), and Rating 5 is 12 responses (50%). All respondents rated either 4 or 5, showing a highly positive evaluation. The mean score is 4.50, reflecting intense overall satisfaction or agreement. The standard deviation is 0.50, indicating low variation among participants' responses (see Figure 6).

3. The Question & Answer session is helpful for learning

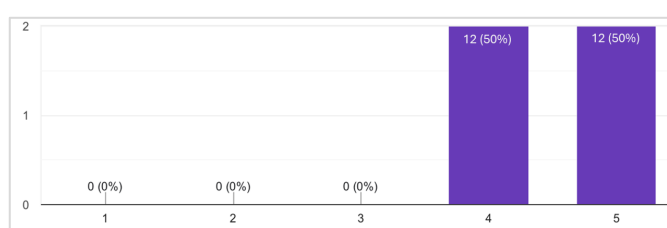


Figure 7. Results of answering Question 3 of the questionnaire after the experiment.

Rating 1-3 is 0 responses, Rating 4 is 12 responses (50%), and Rating 5 is 12 responses (50%). All respondents rated either 4 or 5, showing a highly positive evaluation. The mean score is 4.50, reflecting intense overall satisfaction or agreement. The standard deviation is 0.50, indicating low variation among participants' responses (see Figure 7).

4. Do you think the instructions are clear?

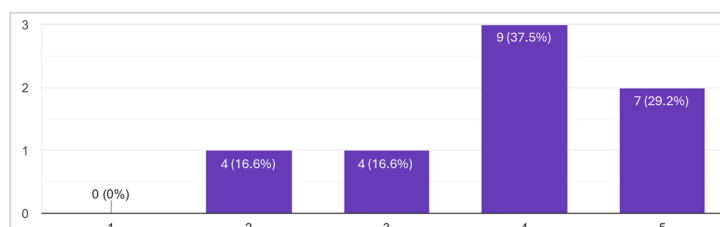


Fig. 8. Results of answering Question 4 of the questionnaire after the experiment.

Rating 1 is 0, Rating 2 is 4 responses (16.6%), Rating 3 is 4 responses (16.6%), Rating 4 is 9 responses (37.5%), and Rating 5 is 7 responses (29.2%). The mean value of the dataset is approximately 3.79. The standard deviation is approximately 1.04, which reflects a moderate spread of the data around the mean (see Figure 8).

5. Do you think Image Narrator is easy to use?

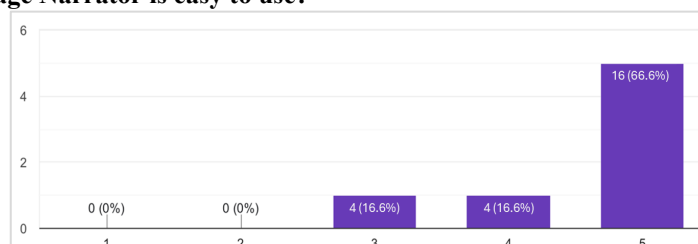


Figure 9. Results of answering Question 5 of the questionnaire after the experiment.

Rating 1-2 are 0, Rating 3 is 4 responses (16.6%), Rating 4 is 4 response2 (16.6%), and Rating 5 is 16 responses (66.6%). The mean value is approximately 4.5. This suggests that most participants selected higher response options, learning towards option 5. The standard deviation is approximately 0.76 (see Figure 9).

5 Conclusion

This study presents *Image Narrator*, an AI-enabled accessibility system that assists visually impaired learners in understanding video-based instructional content. By integrating a Chrome extension, a web platform, and generative AI-powered visual recognition, the system extracts visual information from online tutorials and converts it into accessible textual or auditory formats. User testing at CNIB showed consistently high satisfaction with the system's functions, demonstrating that AI visual narration can significantly reduce learning barriers, support independent learning, and broaden access to online STEM educational materials for visually impaired students.

Referencias

- [1] Ahmed, M.R., Naveed, M.A. (2020). Information accessibility for visually impaired students. *Pakistan Journal of Information Management & Libraries*, 21(1), 25–38.
- [2] ChromeVox Screen Reader Install and Setup – Web Accessibility for Developers. <https://pressbooks.library.torontomu.ca/wafd/chapter/chromevox-screen-reader-install-and-setup/>
- [3] Hakim, M.I., Md Yusof, H. (2020). A mechatronics approach to develop STEM accessibility tools for visually impaired students. In: *Proceedings of the 6th International Conference on Robot Intelligence Technology and Applications*, 171–183.
- [4] JAWS. <https://www.freedomscientific.com/products/software/jaws/>
- [5] NV Access. <https://www.nvaccess.org/>
- [6] Rule, A.C., Stefanich, G.P., Haselhuhn, C.W., Peiffer, B. (2009). A Working Conference on Students with Disabilities in STEM Coursework and Careers. ERIC.
- [7] Rule, A.C., Stefanich, G.P. (2012). Using a thinking skills system to guide discussions during a conference on students with disabilities pursuing STEM fields. *Curriculum & Instruction Faculty Publications*, 17.
- [8] The Vision of Children Foundation (2018). The rise of artificial intelligence for the visually impaired. <https://www.visionofchildren.org/voc-blog/>
- [9] Wang, J., Wang, S., Zhang, Y. (2022). Artificial intelligence for visually impaired. *Neural Computing and Applications*, 34, 7905–7924.
- [10] Wang, J., Wang, S., Zhang, Y. (2023). Artificial intelligence for visually impaired. *Displays*, 77, 102391.
- [11] Xu, W., Ouyang, F. (2022). The application of AI technologies in STEM education: A systematic review from 2011 to 2021. *International Journal of STEM Education*, 9, 59.