EasyWiki: A Creative Support Tool For Education in Underrepresented Languages

David Moerdijk
Leiden University
s3178749

***Abstract*—This report presents *EasyWiki*, a creative support tool designed to generate reliable and easily digestible educational videos in underrepresented languages. The system leverages recent advancements in generative AI across multiple modalities in text generation (llm's) and voice synthesis (GANs and transformer architectures [27]), while also making use of more tried and true methods for algorithmic video editing. The system makes use of Wikipedia's catalog of more than 60 million articles across many topics. EasyWiki addresses the gap in educational resources for less-represented languages across the globe. With more than 55% of people across the globe not being able to speak one of the six UN-recognized languages [7]. As educational resources primarily target speakers of the six UN-recognized languages, those who do not speak these languages face a significant barrier to learning. The tool makes it possible for teachers, creatives, and pupils to automatically create videos of 1-10 minutes, which then can be used as easily digestible educational content. Key features of the system include a fully automated two-step workflow of asset generation and asset combination. Results qualitatively demonstrate the system's ability to produce educational videos in languages with diverse scripts, with tests including Amharic and Indonesian. While successful in its aims, limitations such as voice quality and presentation aesthetics remain, leaving possibilities for improvements. EasyWiki exemplifies how technology can moore easily democratize access to information than has ever been possible, bridging language barriers worldwide and enhancing education globally.**

I. Introduction

For the last few years, the field of Generative AI has undergone transformative changes. Across multiple modalities, models have evolved from being intriguing proofs-of-concept to practical tools with many real world applications [1]. Many have tried: books have been published purely written by LLM's [6], fictional movie trailers have been published where all shots were purely generated by Generative AI models [16], and many more instances. However, even the current state-of-the-art models have many limitations. These models still miss many intuitions that experienced humans in a field have. One of the biggest problems is the concept of hallucination. This can be defined as *a plausible but false or misleading response generated by an artificial intelligence algorithm* [21]. The goal of this project is to use the newest and best advancements in Generative AI models to create forms of content in a reliable way, utilizing methods that minimize the hallucinations in the content to improve educational accessability. As Baron (2021) has described, educational content is available for free to many with an internet connection. Some problems that still persist herein are that content is often available, but not in the format that the person searching for education on a topic is most comfortable with.

> The choice of materials – print or digital, audio or video – can affect learning outcomes. Beyond cost and convenience, digital media have their advantages over traditional printed texts. A skilled audiobook narrator can keep listeners riveted, and watching a video on climate change can have a bigger impact than reading about the subject.
> - Naomi Baron, 2022 [4]

For example, take a child speaking a globally underrepresented language, such as e.g. Amharic, wanting to learn about the history of the great wall of china. An English-speaking child would have no problem finding an abundance of the right content for that on youtube. The Ethiopian child would likely try, but find little, and possibly give up. English speakers often take for granted the ease with which one can easily learn all there is to know about anything they want in easily digestible videos made about the subject. However, finding equally easily digestible educational content in a language that only x% of the world speaks (Amharic, in this case) [5][19] would be much harder than finding said content in a language that 18.8% percent of the world speaks [10][19] without having to sift through hard-to-digest material. Research has shown that a persons' ability to learn is greatly influenced by their enjoyment of digesting that content [25]. Leveraging short-form videos to create educational content in diverse languages has the potential to significantly enhance learning and enjoyment for the billions of people (more than 4.4 billion to be exact) worldwide who do not speak one of the most widely spoken languages. The six official languages of the United Nations are Arabic, Chinese (Mandarin), English, French, Russian, and Spanish. A majority of the world, 55%, does not speak any of these languages as first or second language [7]. See Figure 1 for a visualization.

This report will look at the challenges in being able to generate - or translate - forms of content in an easily digestible, reliable and enjoyable way for people across the globe.

II. Background Research

The aim of this section is to contextualize the status quo, and concretize the challenges in creating digestible, reliable, and diverse educational content for underrepresented languages.

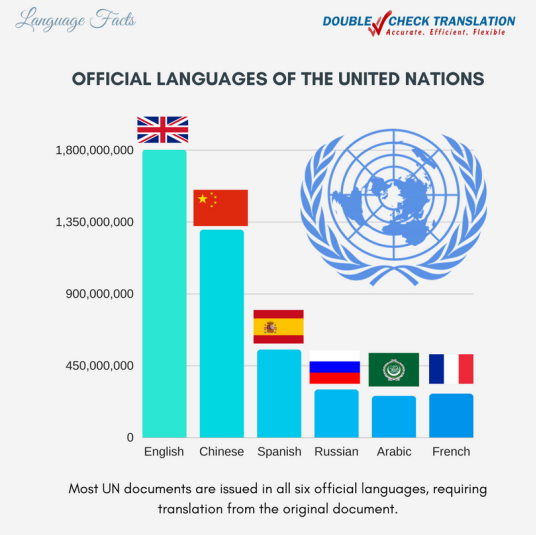

Fig. 1. Amount of speakers per officially recognized United Nations language

A. Similar Existing Systems

Many different systems are already in place that are able to create videos with the help of Generative AI requiring a varying amount of input. Many of these complete video generators are often more commercial, focusing on areas such as sales, marketing, or just general video creation. A shortcoming many of these generators have is a very limited input, perhaps only a short prompt given by the user, such as ”A product ad for coca cola with a christmas theme” [14][17][11]. Other systems have taken an approach less susceptible to hallucination [20]; using extensive input documents [18][23]. Another great example in the sector of education is the commercially succesful language learning app *Duolingo*, which shortly after the release of GPT-4 released *Duolingo Max*, with many Generative AI-powered functionalities [26], such as role playing and answer explanation.

B. Resources and Literature

1) Datasets: As this system leverages top-notch models providing quality that could not be achieved in the scope of this report, datasets have not been necessary to train or even finetune models for the system.

However, one important dataset has been used as an important source for information and antidote to the model's hallucinations; wikipedia [30]. The contents from a wikipedia are gathered through scraping the wikipedia website based on any input article link that is given by the user.

*Note: Wikipedia's content is licensed under Creative Commons license **Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)**, which allows to share and adapt its content, even commercially [8]*

*Note: More information on datasets is referenced in **section VI.** under **Datasets***

2) Text generation: Perhaps the most impactful part of asset generation for the video is the script. The text for the script will be generated fully by a large language model. Many different models exist for this task. Advanced generative AI models like OpenAI's GPT-4, Google's PaLM, and Meta's LLaMA have already shown great potential in generating such content. Of the models most easily available with API calls, OpenAI's GPT-4 performs the highest on many tests [13], while still offering an affordable price tag. It scored 88.70% on MMLU, second in place behind Claude 3.5 Sonnet. Using local models would severely limit the devices the whole system could be run on, and this option is therefore not chosen.

3) Key articles: One important research article that originally sparked the idea for this video generation system is ”Common Tongue: The Impact of Language on Educational Outcomes” by Jain (2017) [15]. It follows one of the linguistically most diverse countries; India. Here, it found that in some regions, schools taught in another language than what was the native spoken language for the pupils, and compared those to children in other regions where they did natively speak the same language as their education language. It found that the mismatched regions had 18.0% lower literacy rates and 20.1% lower college graduation rates. The study also explains how these lower rates were corrected once the correct education language was matched to the pupil's natively spoken languages. This article underpins the goal of this system very well; people learn much better about subjects in their native languages.

Many other articles have found similar phenomena. Chami found in 2024 how the effect is even stronger for *children* coming from a middle-income background [3].

> Findings call for prioritizing research and policy attention to language of instruction and linguistic discordance globally toward achieving basic education for all.

An important disclaimer is that many may bring up the positive effects bilingual schooling often has on children's literacy rates in both languages [2]. This is found to be true by many studies, but I would like to make clear here that that is not the type of situation this video generation tool is made for. Some children might bilingually learn Malay and English for example, as is the case in many Singaporean schools [9]. Those could benefit from the large array of english education tools that are available online. However, this tool is meant for children (and other learners) that do not speak any of the languages that provide such large online education.

The type of content is similar to a system of education that is quickly growing in recent years online, called MOOC (Massive Online Open Course). It is a form of educational content where students can learn for free about a plethora of topics, often done in the format of narrated video's with accompanying animations or images. If the explanation is purely audio-based without visual stimulation, students can struggle much more than with [28].

III. Project Aims

The aim of this project is to create a system that allows teachers, creatives, and pupils to automatically create videos of 1-10 minutes that are easily digestible where the only necessary inputs are a wikipedia article and a target language.

A. Problem Statement

The current global educational system faces a problem of accessible and engaging resources for speakers of underrepresented languages. This goes directly against the United Nations' Sustainable Development Goal SDG-4, which *aims to ensure equitable and inclusive quality education and lifelong learning for all* [22]. The effect of this problem, is that people living in poverty in third world countries face much larger hurdles in escaping from poverty than they should. The problem can be summarized as: *A large problem in global education today is the challenge of finding digestible, reliable, and diverse educational content for underrepresented languages.*

B. Creative Activity Supported & Support for Creativity

The primary activity supported by EasyWiki is the transformation of raw text of a Wikipedia article into rich, multimodal content that engages and educates the viewer. The goal, broader than simply video creation, would be to allow for creators to create complete thematic video series on video sharing platforms such as youtube, needing little input from the creator in terms of details of implementation, leaving more room for (1) deep thought of what viewers want and (2) marketing the content to the correct target groups.

The domain supported will be the creation of educational content, and thereby also the act of education itself.

C. Target Audience

1) Creators: The target audience for the people creating the videos would be people with little time and much experience teaching. A great example would be a teacher using the tool as part of creating a set of resources to watch for the homework.

2) Viewers: While the system is designed to benefit a broad spectrum of viewers, the focus is on young students aged 8-14 who are often underserved by existing educational resources. Detailed choices on how the system will work will largely be based on what would be most beneficial to this group. This does not have to make the generated content useless to watch for older learners.

IV. Implementation

The process of implementing this system requires many different modalities of generative models and other data generation API's to be combined. The way the codebase works is that the process of video generation is divided into two separate steps; *generation* and *combination*.

These modules can be found in the codebase as *backend.generation* and *backend.combination*. Notice how there is no *frontend* module class yet.

1) Generation: Generation refers mainly to the *Generative* models creating assets, but also to the gathering of assets through the use of web scraping and asset gathering modules.

Assets can be generated using many methods, and the type of generation method is defined by the variable *generation_method* in a *config.json* file. Some of those are *stock_video* (gathering stock videos for filler in end result), *website_picture* (for capturing a screenshot of a website), *voice* (for generating the voiceover .mp3 file given a voiceover script), and many more.

2) Combination: Combination also uses the config.json file to combine many different assets gathered in the asset generation/gathering step, combining it into one end result video. Similarly, the workflow for this part of the module takes in the config.json file, which is the main interface through which data is exchanged between the two modules.

Combination is done using the *moviepy* python module, which in turn is written using one of the most well-known programs for algorithmically editing videos; *ffmpeg*.

Combination allows for many different functionalities, the most important of which are adding a background image, adding overlay images on top of that background image, adding subtitles, adding multiple streams of audio, and some more specific functionalities.

A. Type of Project

The type of project is meant to be a fully autonomous system that takes in two simple but extensive inputs: a *wikipedia url*, and a *target language*. Given the huge amount of wikipedia pages available online, as of this report totalling more than 60 million on specific topics.

With wikipedia being one of the most up-to-date, but still very extensive sources of information currently available on the internet, the project is meant to allow for a creative freedom, where the creators are able to create videos on any topic, or even a video series on a theme of topics as part of their teaching curriculum.

B. Resources Used

1) APIs and Web Services:

- **Pexels API:** Used to fetch stock images based on search terms as thought of by the coordinating LLM prompt.
- **Unsplash API:** Alternative for stock images.
- **YouTube API:** Enables downloading videos based on specific search terms. Not used in the main wikipedia workflow anymore, but still useful for other workflows.
- **ElevenLabs API:** Utilized for generating voiceovers (and synchronized subtitles).
- **ScreenshotLayer API:** Captures screenshots of websites for inclusion as assets. Also not used in the main wikipedia workflow anymore, but still useful for other website / product ad-based workflows.

2) Special Python Libraries:

- **moviepy:** Core library for video and audio manipulation, including asset combination and subtitle integration.

Fig. 2. Screenshot of educational video generated using EasyWiki on Justin Trudeau in English

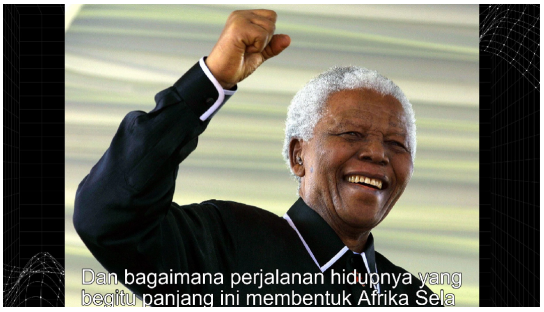

Fig. 3. Screenshot of educational video generated using EasyWiki on Nelson Mandela in Indonesian

Note: basic python libraries like requests and json are not specifically mentioned, as this does not provide any useful extra information.

3) Datasets and Pre-trained Models:

- **Generative AI Models:** GPT-4 (OpenAI) for generating text and translations.
- **Generative AI Models:** ElevenLabs for generating the voiceover.

C. System Architecture

D. Implementation Details

V. Example Outputs

A. English

To allow for some better interpretability for the english reader of this article, here are first provided some english-spoken examples. See an example of such a video created on Justin Trudeau using his wikipedia article in Figure 2 [29].

B. Underrepresented Languages

As ElevenLabs only supports 32 languages officially, see Figure 3 for Indonesian. To showcase EasyWiki's ability to generate videos in many languages and scripts, see Figure 4 for an Amharic video.

For easy access, the videos have also been uploaded to YouTube and are accessible at the links below:

- Trudeau (EN): https://youtu.be/G_E6PTekJVI
- Mandela (ID): https://youtu.be/6U1zOXvTe4g
- Great Wall of China (AM): https://youtu.be/71zXPs1Qpd8

Fig. 4. Screenshot of educational video generated using EasyWiki on The great wall of China in Amharic

VI. Reflection and Future Work

A. Achievements

The finished project achieved its aim by succesfully creating a fully working system for generating easily digestible educational content in many different underrepresented languages. No hallucinations have been found, due to the LLM's strict adherence to information found in the wikipedia articles.

B. Limitations

One of the main goals of *EasyWiki* was for an easily digestible format to be created, starting from a much more raw format of a wikipedia article. Though this has definitely been the case, many improvements are possible to make the videos more pleasurable to watch and more professional. First, the specific voice could be improved on. Though elevenlabs offers a large amount of voices to choose from, this system for now uses a very simple, robotic-sounding voice. This was done largely because of budgetary reasons.

C. Future Directions

An approach that many video generator systems have taken is to add a high quality *persona* to the video, giving a more personal explanation feel. Perhaps the current system could benefit from such a presentor in the video, although it might give the generated videos a more robotic or impersonal feel. Though the project has served as a fully working tool, rather than a simple proof-of-concept, it can definitely serve as one in the context that this method could work for many other format-to-format systems. Perhaps in the future Generative AI could be used to translate other information formats such as books into easily digestible video's. A great example would be for a teacher to be able to translate an educational book's chapters into easily digestible informational videos. The format of informational video's has clearly been proven to provide a lot of value for pupils in high school and college. An interesting project by Google, which was also a great part of the inspiration for the current project, could inspire other end-formats for similar educational tools to this one. This Google tool is called NotebookLM [12], and allows a student to upload any array of documents into the tool, after which a two-person podcast will be generated, also allowing for easily digestible content.

1) Datasets: Write based on subsectionDatasets

2) Extra: An important requirement that has been engineered for this system is to be able to support as many languages as possible. Simple fonts like *Arial* were thus not possible to be used, as only the Latin script was supported using that font. Therefore, an open-source font, released by GNU, was used. This font is very simple, but does all scripts than can realistically be expected. In the standard scripts; Latin, Cyrillic, Arabic, Ethiopian, and many more. In the ancient scripts; Old Italic, Coptic, Phoenician, Runic, and many more. The font pack could even be useful for scientific scripts such as the phonetic alphabet, braille, and complex mathematical notations [24].

For easy digestability, fonts should also be visually appealing. The current font is very simple, and not very appealing. In future versions, a more visually appealing font could be used, with perhaps logic that chooses a font for each alphabet manually.

3) Customizability: Beyond simple video creation, the system could encourage creative storytelling by enabling users to tailor the tone, style and visual elements shown in these videos. This would allow for a less frustrating experience for the creator, and a more captivating experience for the viewer.

REFERENCES

APPENDIX

REFERENCES

[1] Safe generative ai. https://neurips.cc/virtual/2024/workshop/84705 (2024), workshop at the 38th Conference on Neural Information Processing Systems (NeurIPS 2024)

[2] Aguilar, G., Surrain, S.: The effects of native language instruction on the language and literacy achievement of prek-6th language-minority students. In: Presented at the Harvard Graduate School of Education. Cambridge, MA (2017), https://scholar.harvard.edu/files/sarahsurrain/files/sssr_poster_7.pdf, this study reviews the effects of bilingual education on English and Spanish literacy outcomes for language-minority students.

[3] Alkateb-Chami, M.: Learning poverty when schools do not teach in children's home language: A comparative perspective. International Journal of Educational Development **105**, 102975 (2024). https://doi.org/https://doi.org/10.1016/j.ijedudev.2023.102975, https://www.sciencedirect.com/science/article/pii/S0738059323002511

[4] Baron, N.S.: Enhancing learning from digital, audio and video resources. Times Higher Education (May 2022), https://www.timeshighereducation.com/campus/enhancing-learning-digital-audio-and-video-resources, accessed: 2025-01-22

[5] Basha, S.J., Veeraiah, D., Charan, B.V., Joyce Yeddu, W.S., Babu, D.G.: Detection and comparative analysis of handwritten words of amharic language to english using cnn-based frameworks pp. 422–427 (2023). https://doi.org/10.1109/ICICT57646.2023.10134103

[6] Bosker, B.: How to author over 1 million books. HuffPost (February 2013), https://www.huffpost.com/entry/philip-parker-books_n_2648820, retrieved 2025-01-22

[7] Central Intelligence Agency: World factbook - languages (2025), https://www.indexmundi.com/world/languages.html, accessed: 2025-01-21

[8] contributors, W.: Text of the creative commons attribution-sharealike 4.0 international license (2024), https://en.wikipedia.org/wiki/Wikipedia:Text_of_the_Creative_Commons_Attribution-ShareAlike_4.0_International_License, accessed: 2025-01-22

[9] Dixon, L.: The bilingual education policy in singapore: Implications for second language acquisition (01 2003)

[10] Eberhard, D.M., Simons, G.F., Fennig, C.D.: Ethnologue: Languages of the World. SIL International, Dallas, Texas, 25th edn. (February 2022), https://www.ethnologue.com, iSBN 978-1-55671-461-6

[11] FlexClip: Flexclip ai video generator (2025), https://www.flexclip.com/tools/ai-text-to-video/, accessed: 2025-01-21

[12] Google: Notebooklm. https://notebooklm.google.com/ (2023), aI-powered research assistant by Google.

[13] II, S.M.W.: Llm benchmarks (July 2024), https://klu.ai/glossary/llm-benchmarks, klu.ai Glossary, Co-Founder / CEO

[14] InVideo: Invideo ai video generator (2025), https://invideo.io/make/ai-video-generator/, accessed: 2025-01-21

[15] Jain, T.: Common tongue: The impact of language on educational outcomes. The Journal of Economic History **77**(2), 473–510 (2017). https://doi.org/10.1017/S0022050717000481

[16] Kantor, J.H.: The 14 best ai-generated fake movie trailers, ranked. Looper (October 21 2024), https://www.looper.com/1686162/best-ai-generated-fake-movie-trailer-ranked/, accessed: 2025-01-22

[17] Kapwing: Kapwing ai text to video tool (2025), https://www.kapwing.com/ai/text-to-video, accessed: 2025-01-21

[18] Lumen5: Lumen5 - video creation platform for transforming text into video (2023), https://www.lumen5.com, accessed: 2023-01-21

[19] Lutz, W., KC, S.: World population growth: A once and future global concern. World **4**(4), 43 (2023)

[20] Maynez, J., Narayan, S., Bohnet, B., McDonald, R.: On faithfulness and factuality in abstractive summarization. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 1906–1919. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.acl-main.173, https://aclanthology.org/2020.acl-main.173/

[21] Merriam-Webster: hallucination. https://www.merriam-webster.com/dictionary/hacker, accessed: 2025-01-21

[22] Nations, U.: The 17 sustainable development goals (sdgs) (2015), https://sdgs.un.org/goals, accessed: 2025-01-22

[23] Pictory: Pictory - video creation made easy with ai (2023), https://pictory.ai, accessed: 2023-01-21

[24] Project, G.: Gnu freefont. https://www.gnu.org/software/freefont/ (2025), a free family of scalable outline fonts, Unicode-encoded, suitable for general use on computers and desktop publishing, available in TrueType and OpenType formats.

[25] Song, Y.: Assessing the interactions between learning enjoyment, motivation, burnout, and grit in efl students: a mixed-methods approach. BMC Psychology **12**(1) (Dec 2024). https://doi.org/10.1186/s40359-024-02303-6, http://dx.doi.org/10.1186/s40359-024-02303-6

[26] Team, D.: Introducing duolingo max, a learning experience powered by gpt-4 (March 14 2023), https://blog.duolingo.com/duolingo-max/, accessed: 2025-01-21

[27] Team, V.: What is elevenlabs + how to use it [2024 tutorial] (July 2024), https://www.voiceflow.com/articles/elevenlabs, published on Voiceflow's blog and provides detailed tutorials and insights into AI voice technologies.

[28] Uchidiuno, J., Hammer, J., Yarzebinski, E., Koedinger, K.R., Ogan, A.: Characterizing ell students' behavior during mooc videos using content type. In: Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale. L@S 2017, ACM (Apr 2017). https://doi.org/10.1145/3051457.3053981, http://dx.doi.org/10.1145/3051457.3053981

[29] Wikipedia contributors: Justin trudeau — wikipedia (2025), https://en.wikipedia.org/wiki/Justin_Trudeau, accessed: 2025-01-21

[30] Wikipedia contributors: Wikipedia, the free encyclopedia (2025), https://www.wikipedia.org, accessed: 2025-01-22