

Article

Not peer-reviewed version

Generative Text-to-Video AI for Linguistically Responsive Lesson Presentation: Towards a Multimodal Intelligent Classroom Assistant

[matthew nyaaba](#)* and SungEun Min

Posted Date: 22 November 2024

doi: 10.20944/preprints202411.1743.v1

Keywords: Text-to-video; linguistically responsive teaching; Generative AI; Bilingual teaching; AI; Multimodal Intelligent



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Generative Text-to-Video AI for Linguistically Responsive Lesson Presentation: Towards a Multimodal Intelligent Classroom Assistant

Matthew Nyaaba ^{1,*} and SungEun Min ²

¹ AI4STEM Education Center and Department of Educational Theory and Practice, Athens, Georgia, USA

² Elementary, Middle Level, Library and Technologies Education Department (EMLT), Kutztown University of Pennsylvania

* Correspondence: matthew.nyaaba@uga.edu

Abstract: In this study, we explore the potential of Generative Text-to-Video AI (GenT2V AI) to function as a *Multimodal Intelligent Classroom Assistant* (MICA). This tool is capable of presenting lessons with voice-overs, instructor avatars, language translations, and culturally relevant animations to support linguistically responsive teaching in diverse classrooms. Using a GenT2V AI tool, we converted a lesson presentation targeting a diverse group of pre-service teachers, consisting of English-speaking students from the United States and Spanish-speaking students from Mexico and Chile. The conversion employed several functionalities, including generating English audio in a U.S. accent, translating captions into Spanish, creating reflective animations for the content on "Education in Ghana," and representing the presenter (instructor) as an avatar. We analyze the output based on these functionalities. Our findings indicate that GenT2V AI has the potential to enable English as a Second Language (ESL) instructors to deliver content effectively. Furthermore, the GenT2V AI can enhance lesson accessibility for ESL students where English is the medium of instruction. The interactive aspects of GenT2V AI could also foster greater student engagement during lessons. However, we identified several limitations: languages and accents underrepresentation, premium cost subscriptions as a barrier to some teachers. Also, advanced pedagogical skills are required to effectively integrate GenT2V AI presentations with interactive elements, such as questioning and pausing for discussions. We recommend the incorporation of culturally diverse languages, accents, visuals, and avatars, as well as sign language overlays, to enhance accessibility. There should be empirical studies to gather insights from teachers and students.

Keywords: Text-to-video; linguistically responsive teaching; Generative AI; Bilingual teaching; AI; Multimodal Intelligent

1. Introduction

Imagine having a Generative AI tool that presents your lesson in a multimodality approach, with a voice-over, an avatar of you, translation of the content in diverse languages, and generating relevant background content of your lesson with animation—a Multimodal Intelligent Classroom Assistant (MICA). Generative text-to-video AI (GenT2V AI) conversion in education is a cutting-edge approach. GenT2V AI leverages advancements in GenAI to potentially generate multimodal output to enhance teaching methodologies, especially in linguistically diverse classrooms [1]. As the global classroom becomes increasingly multicultural, educators face the challenge of addressing varied linguistic needs to ensure equitable learning opportunities for all students. English Language Learners (ELLs), students who are acquiring English as a new or additional language, make up one of the fastest-growing populations in U.S. PreK-12 classrooms [2]. Between 2011 and 2021, the number of ELLs increased from 4.6 million to 5.3 million. In 13 states, including Texas, California, and New Mexico, ELs comprised 10 percent or more of the public-school student population. Moreover, more than 400 different languages are spoken in U.S. public schools [2]. Among these, approximately 4 million students speak Spanish as their home

language, while about 130,000 speak Arabic. These statistics highlight the increasing importance of culturally and linguistically responsive teaching for ELLs.

However, research shows that teachers are not prepared to teach this growing group of ELs in the diverse settings like the United States [3–8]. When teachers have ELs in their classrooms who speak languages other than English, monolingual teachers often feel uncertain about how to proceed because they lack the knowledge and skills needed to effectively teach this linguistically diverse group of students [8]. There is a need to create an ongoing effort to develop linguistically responsive teaching practices that can better address the challenges faced by educators and ELLs.

The influence of different media formats on students' learning outcomes has been a subject of ongoing research, with studies like that of Tarchi et al. [9] demonstrating substantial equivalence between text and video formats in immediate comprehension. Further studies demonstrate the potential of multiple AI tools, such as HeyGen, ElevenLabs, DeepL, and Adobe Character Animator, for producing multilingual and multimodal content, primarily in MOOCs and higher education [10]. Moreover, Dao et al. [11] utilized text-to-speech and speech-driven-face technologies to automatically generate video lectures with an instructor's voice and face, where their findings show that this approach enhances learner engagement and flexibility. While GenT2V AI tools have been successful, they often lack the simplicity required in educational content, which is critical for making complex materials accessible to learners with diverse linguistic or cognitive abilities [10,11]. Recommendations have been made for the development of frameworks to ensure high-quality multilingual educational content.

This study builds on these recommendations by employing the GenT2V AI tool, VEED, a GPT within the OpenAI platform, as a prototype for a Multimodal Intelligent Classroom Assistant (MICA). This GenT2V AI tool integrates external web platforms to create content with captions, voiceovers, and animations, offering a streamlined approach to classroom instruction. This study explored the capabilities of this GenT2V AI and its potential to enable English as a Second Language instructors to deliver content in multiple languages and accents, making lesson presentations more accessible for English Language Learners in schools where English is the medium of instruction, such as in the United States.

2. Generative Text-to-Video AI in Education

Generative Text-to-video models have significantly advanced multimedia content generation by transforming textual descriptions into dynamic video sequences. Platforms like Sora, as explored by Adetayo et al. [12], illustrate the promise of AI-powered text-to-video generation in education. Sora's ability to transform text into dynamic video content has proven useful in creating interactive, personalized, and accessible educational materials. The potential for immersive storytelling and educational gamification also points to the broader application of text-to-video models in making education more engaging and inclusive [12]. These models leverage deep learning techniques that combine natural language processing and computer vision to generate videos accurately reflecting the narrative conveyed in the text. Advanced methods like attention mechanisms enhance the models' focus on critical details, improving the accuracy of video generation [13]. Evaluating these models presents challenges due to the limitations of automatic metrics, leading to a reliance on manual evaluations [13]. Innovations like Text-driven Video Prediction (TVP) focus on the causal relationship between text and motion for more coherent video generation [14].

A recent study discusses five text-to-video models: Tune-a-Video, VideoFusion, Text-To-Video Synthesis, Text2Video-Zero, and Aphantasia, evaluating their generated video quality against various metrics [15]. These models rely heavily on the quality of video-text pairs, yet current datasets suffer from major deficiencies, including low temporal consistency, substandard captions, poor video quality, and imbalanced data distributions [16]. Efforts to address these challenges are exemplified by VidGen-1M,

a newly developed dataset specifically designed to improve the performance of text-to-video models through a coarse-to-fine curation strategy [16]. However, there remains a gap in how these technological advancements can be applied to education, particularly when it comes to accommodating multilingual learners with varying accents and linguistic backgrounds [16]. Integrating advanced datasets like VidGen-1M with tools such as VEED and Rendora, which support multilingual features and diverse accents, offers a potential solution. Such integration can facilitate the creation of high-quality, culturally relevant, and linguistically diverse educational content, advancing the role of text-to-video models in education [12,16].

3. Linguistically Responsive Teaching

This study situates itself within linguistically responsive teaching while examining the tenets of culturally responsive pedagogy, emphasizing the need to view culture as an asset-based concept and exploring the efficacy of Generative AI videos from recent studies. Lucas and Villegas [17,18] studied teachers who exemplified responsive teaching practices and introduced the Linguistically Responsive Teaching (LRT) framework. They expanded the concept of cultural responsiveness to include linguistic responsiveness to address the specific needs of English Language Learners (ELLs) in the U.S., whose first language is not English. Scholars deliberately use the term “linguistically responsive teaching” to differentiate between students of color (e.g., racial identities) and multilingual learners (e.g., linguistic identities). While both groups represent historically underrepresented populations in the U.S., grouping them together can be problematic, as it overlooks the distinct needs of each group [8].

The key tenet of linguistically responsive teaching is that teachers should view multilingual learners' primary language as an asset. Given that the home language of a multilingual learner will play a role in acquiring English language and literacy, it should not be discouraged for the learner to use it in the classroom [19]. Linguistically responsive teaching posits that multilingual learners' language and literacy knowledge, which is developed in their first language (e.g., Chinese), can transfer to their learning of English, a second language [8,17].

4. Dimensions of Culturally and Linguistically Responsive Instruction

In the context of culturally and linguistically responsive instruction, Montenegro and Jankowski [20] emphasize three key dimensions that should shape equitable assessment practices: explicit elements (such as language, cultural practices, and heritage), implicit elements (like beliefs, values, and shared experiences), and cognitive elements (how lived experiences influence learning, communication, and knowledge expression). Language, as a fundamental aspect, the explicit dimension, transcends all three, deeply influencing students' identities and learning processes. With tools like VEED AI, which can convert text across diverse linguistic backgrounds, educators can place students' linguistic identities at the center of instruction, aligning with culturally responsive practices that promote equity.

Additionally, Lucas and Villegas [17,18] proposed a pedagogical framework of culturally and linguistically responsive instruction that comprises both orientation and knowledge/skills aspects to accommodate the diverse learning needs of multilingual learners. They conceptualize four essential knowledge and skill dimensions. These include strategies for learning about the linguistic and cultural backgrounds of students, knowledge of and ability to apply key principles of second language learning, identifying task-specific language demands, and strategies for scaffolding instruction.

5. Asset-Based Theory

Asset-based education focuses on recognizing and leveraging the strengths, experiences, and cultural knowledge that students bring to the learning environment [21,22]. Rather than focusing on perceived deficits, this approach views students' diverse linguistic, cultural, and community backgrounds

as valuable resources that enhance learning and engagement. Asset-based practices tap into students' cultural wealth, such as their home languages, community networks, and lived experiences, to make education more inclusive and equitable [23]. GenT2V AI's capability to convert lessons into multiple languages through both audio and captions aligns perfectly with asset-based education by supporting linguistically responsive teaching. GenT2V AI allows educators to present educational content in students' native languages, fostering greater accessibility and engagement [24,25]. This feature ensures that students from diverse linguistic backgrounds are able to connect with the content in a way that honors their cultural identity, making learning more meaningful and personalized. Additionally, GenT2V AI facilitates translanguaging practices, where students can fluidly use their home languages alongside academic content, thus reinforcing their understanding and allowing them to fully engage with the material [26]. It is believed that GenT2V AI may encourage educators to implement asset-based strategies that recognize and validate students' linguistic assets, promoting diversity, equity, and inclusion in education for all learners.

6. Approach

Using an exploratory approach, we analyze the functionalities of GenT2V AI, using the VEED GPT as a case study. We focused on how GenT2V AI can support instructors by enabling multilingual video content that enhances the accessibility of lessons for students from different cultural backgrounds. Specifically, we utilized GenT2V AI to convert a PowerPoint lecture on "Education in Ghana," targeting diverse pre-service teachers from Spanish-speaking and U.S. English-speaking backgrounds. While the audio was in U.S.-accented English, the captions were provided in Spanish. This approach aimed to address communication differences that posed barriers to learning.

6.1. The VEED GPT

The VEED.io GPT Text-to-Video tool combines OpenAI's generative capabilities with VEED.io's video processing features to create a powerful solution for multimedia content. It offers automated transcription for precise multilingual text, subtitling and captioning with customization options, and video content analysis to highlight key moments and generate summaries. The tool supports real-time translation, interactive editing, and content summarization, making it ideal for bridging language gaps and enhancing accessibility. With seamless integration via APIs and multimodal accessibility, it processes audio, video, and text to deliver outputs such as fully captioned videos, concise transcripts, or summaries, offering an engaging, inclusive approach that can be utilized in classrooms and beyond.

6.2. Exploratory Analysis of GenT2V AI GPT Model's Functionality

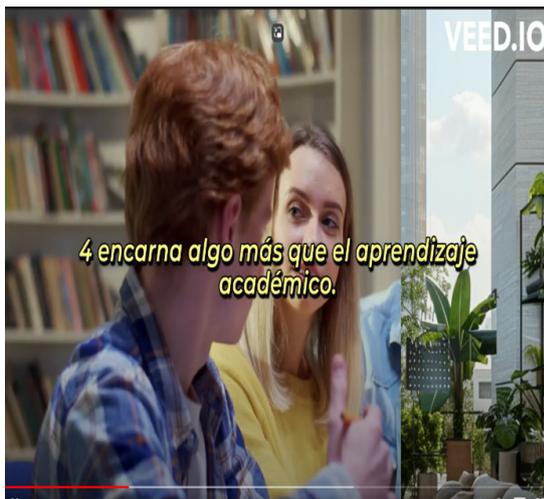
To analyze VEED's features, we performed a demo video on the lesson "Education in Ghana," targeting both English and Spanish-speaking students in a K-8 classroom. The demo focused on converting the lecture content to suit Mexican students with a Spanish language background while adjusting the spoken accent to an American accent, as most students were from the U.S. We incorporated Spanish captions to enhance comprehension and accessibility for the non-ESL students. Additionally, we employed one of VEED's avatar-based presentation features, which allowed for a visually engaging and consistent delivery of the content (see Figure 1). Furthermore, the animations and video content in the demo were automatically generated by VEED to match the lesson's context. VEED's AI-driven tools effectively analyzed the content of the lesson on "Education in Ghana" and produced relevant visual elements that enhanced the overall presentation.



(a) Image 1



(b) Image 2



(c) Image 3



(d) Image 4

Figure 1. Screenshot of VEED avatar delivering a lecture on Education in Ghana, accompanied by Spanish captions and an American-accented voiceover

Using an exploratory approach, we analyze the functionalities of GenT2V AI, using the VEED GPT as a case study. We focused on how GenT2V AI can support instructors by enabling multilingual video content that enhances the accessibility of lessons for students from different cultural backgrounds. Specifically, we utilized GenT2V AI to convert a PowerPoint lecture on “Education in Ghana,” targeting diverse pre-service teachers from Spanish-speaking and U.S. English-speaking backgrounds. While the audio was in U.S.-accented English, the captions were provided in Spanish. This approach aimed to address communication differences that posed barriers to learning.

In the excerpts provided, the demo video used VEED’s standard avatar feature to deliver the lecture on “Education in Ghana,” accompanied by Spanish captions and an American accent. While GenT2V AI’s premium version offers more advanced customization options, such as allowing instructors to use their own pictures in the avatar or personalize the video with uploaded images or videos, these features

were not utilized in this case. In the premium version, instructors could further enhance the video's engagement by incorporating their likeness into the avatar, making the presentation more personal and relatable for the students. Additionally, the ability to upload customized content, such as specific images or videos related to the lesson, could provide a more tailored and immersive learning experience. However, in this demo, the video relied on GenT2V AI's pre-existing avatars and automatically generated visuals, which still delivered an effective, though less personalized, educational experience.

7. Analysis of GenT2V AI as a Tool for Classroom Lectures

This analysis explores the potential of GenT2V AI (VEED) to enhance accessibility and engagement in multilingual and multicultural classrooms through its multimodal capabilities. It evaluates key features such as multilingual audio, captions, and animations while addressing limitations in cultural adaptability, linguistic diversity, and cost.

7.1. Potentials of GenT2V AI as Multimodal Intelligence

GenT2V AI offers several potential benefits in educational settings, particularly for supporting multilingual and multicultural classrooms. One of its main advantages is enhanced accessibility, as multimodal intelligent features allow content to be delivered in different languages, with the option to provide audio in one language and captions in another. This helps break down language barriers and makes lessons more accessible for students with varying language proficiencies, especially for ESL students. It has the potential to streamline content creation, enabling instructors to efficiently convert written lessons into video format by automating tasks like script generation, speech synthesis, and subtitle creation, saving educators significant time. Furthermore, the use of avatars and animations adds a visually engaging element to lessons, potentially increasing student involvement. While there are limitations to avatar diversity, the ability to upload customized avatars with premium versions is an added advantage that could make lessons feel more connected to the instructor. The combination of multilingual support and customizable accents allows instructors, especially non-English-speaking instructors, to cater to the cultural and linguistic needs of their students, promoting a more inclusive and engaging learning environment.

7.1.1. Content Quality and Accuracy

The content generated by GenT2V AI was engaging due to its animated visuals, a feature that holds the potential to capture students' attention. However, when using the free version, the animations and images failed to reflect local cultural contexts, such as those specific to Ghana. Instead, the content was heavily Westernized, making it less relatable for students in African classrooms. While the captions were translated correctly, the generated content did not align closely with educational standards or regional cultural nuances, limiting its effectiveness. This could be a result of the limitations of the free version used.

Furthermore, the voice generated by the avatar was average in pacing and naturalness, though it retained a robotic quality. The American accent of the voice-over was clear but lacked options for regional accents or tonal variations that would better cater to a diverse student demographic. While GenT2V AI offers promising features for multimodal classroom assistants for lecture presentations, its lack of cultural sensitivity and adaptability to local standards remains a significant limitation.

7.1.2. Accessibility Features

The captioning feature was a strong point of the tool. The captions were accurate, synchronized with the audio, and presented in a clear format. This ensures that students who rely on text for comprehension or who have hearing impairments can follow the presentation effectively. The voice-

over, while understandable, provided limited customization options for accents, languages, and tones, which is crucial in multicultural and multilingual classrooms. Additionally, while GenT2V AI offers text alternatives such as transcripts, the free version did not support enhancements like sign language overlays, which would improve accessibility for visually impaired students. The positioning of the avatar in the video could also be adjusted, suggesting that further usability enhancements are needed to optimize student engagement.

7.1.3. Engagement and User Experience

The presentation videos generated with GenT2V AI appeared engaging to the authors, suggesting a potential for increased student attentiveness and retention during lessons. The combination of visuals and audio was effective for maintaining interest, an important feature for facilitating comprehension in diverse classrooms. However, the lack of support for multiple languages and accents limited the tool's flexibility for broader use cases. The avatar's customization was restricted in the free version, meaning access to more relatable and culturally appropriate avatars required a premium subscription. This subscription adds a cost-intensive barrier for average classrooms, particularly in under-resourced educational settings. While the free version offers functional basics, the premium options necessary for cultural and linguistic relevance may not be feasible for all educators.

7.2. Limitations of the GenT2V AI

There are notable partialities in GenT2V AI when it comes to the representation of languages, accents, and cultural identities in educational settings, which can have a negative impact on learners from marginalized communities. Not all languages and accents are adequately represented or supported in the tool, leading to a lack of inclusivity. For students whose languages or accents are not embedded in the system, this may result in them feeling underrepresented or even inferior. For example, certain minority languages may not be available, or accents may not accurately reflect regional varieties, making the learning material less relatable and more difficult for these students to engage with.

Another critical issue is the digital divide, which is exacerbated by the cost involved in accessing premium features of GenT2V AI. While the free version offers basic functionality, advanced customization options, such as a wider range of accents, languages, and avatars, are locked behind a paywall. This creates an unequal learning environment, where students in poor or rural communities may lack access to the full range of features needed to create engaging and inclusive educational content. The financial barriers can limit the tool's potential to assist schools and learners in underserved areas, widening the gap between wealthier, well-resourced institutions and those in marginalized settings.

Furthermore, the available GenT2V AI avatars within the tool largely reflect white, Westernized personalities, leading to an underrepresentation of diverse racial and cultural identities. This lack of avatar diversity can alienate students of color, making them feel excluded or overlooked in the learning process. The limited representation of non-white characters perpetuates a biased educational experience, as students may struggle to see themselves reflected in the digital materials used for their instruction. This reinforces the need for broader representation to ensure that all students feel equally valued and included in the learning environment.

Table 1. Summary of Analysis of GenT2V AI in Supporting Linguistically Responsive Teaching

Feature	Description	Connection to LRT
Multilingual Accessibility	Provides multilingual audio and captions to support ESL learners and diverse linguistic needs.	Promotes inclusiveness in multicultural classrooms.
Contextualized Content Creation	Automates content animations, speech synthesis, and subtitle generation to reduce educator workload.	Enables efficient creation of inclusive and engaging materials.
Engagement and Retention	Uses animated avatars and visuals to capture attention and improve comprehension.	Fosters understanding and retention among diverse learners.
Cultural Adaptability	Offers customizable options like accents and avatars to reflect classroom diversity.	Enhances representation and inclusivity in educational content.
Accessibility Features	Supports students with disabilities through captions and text alternatives, ensuring accessibility.	Provides basic accessibility to all users, promoting equity.
Differentiating Instruction for Language Development	Multimodal features (audio, captions, animations).	Provides varied ways for students to access content, supporting diverse language proficiencies.
Limitations of the GenT2V AI	Partial representation of languages, accents, and cultural identities; cost-related barriers to accessing premium features; underrepresentation of diverse avatars.	Impacts marginalized communities negatively, creating unequal learning environments and reinforcing biases.

8. Discussion

GenT2V AI and its captioning features ensure that content can be made accessible in various languages. The use of video lessons with multilingual support presents significant potential in PreK-12 education, particularly in countries like the U.S., where there is a diverse student population with varied linguistic backgrounds. Research indicates that counties like Gwinnett County in the USA have approximately 80% of students who are bilingual, speaking languages other than English, with English as their second language [27]. This innovative tool could support teachers who may lack the capacity to address all the linguistic differences in their classrooms, which can affect the foundational learning of students in their early years [28,29].

GenT2V AI can foster linguistically responsive teaching and culturally responsive pedagogy. The content generated by the tool is based on the topics under instruction and can incorporate culturally relevant aspects into lessons [30,31]. These lessons can be modified by the instructor to better reflect the cultural diversity of the students, thereby promoting inclusivity and cultural awareness in the classroom. Additionally, the tool has the potential to challenge assimilationist discourse and monoglossic ideologies that often dominate educational settings. By supporting multiple languages and dialects, GenT2V AI encourages pluralism and heteroglossic ideologies, recognizing and valuing the linguistic diversity of students [32,33]. This allows for a more inclusive approach to education, where multiple languages and cultural backgrounds are acknowledged as assets rather than barriers.

GenT2V AI has the potential to serve as a smart tool for teacher professional development. It can help educators learn more about their students' cultural and linguistic backgrounds and assist in interpreting how these backgrounds impact learning. By using GenT2V AI, teachers can develop a deeper understanding of culturally responsive teaching practices and become more adept at incorporating these

practices into their instruction [34,35]. Another positive impact of GenT2V AI is its ability to boost the confidence of non-English-speaking instructors. Research, such as Ghafar [36], has shown that educators teaching in regions where English is the dominant second language face challenges, and GenT2V AI tools could help address these challenges. Furthermore, this study can serve as a foundational exploration for future research on using GenT2V AI, including OpenAI's Sora, for video generation in education [37]. Analyzing features of the GenT2V AI tool provides valuable insights into how it can effectively enhance multimodal teaching in diverse linguistic environments.

Regardless of this potential, GenT2V AI does have limitations. The number of languages and accents it can accommodate may exclude some students, particularly those whose languages or accents are not represented, making them feel left out. This confirms findings that show GenT2V AI faces several challenges, including the realism of instructor voice and face generation, addressing trust concerns to support real-time speech-driven-face capabilities for more interactive and engaging learning experiences [11]. Additionally, teachers need technical skills to effectively manage video-based lectures in in-person classes, ensuring smooth integration of questioning, pausing, and other interactive components. While GenT2V AI offers a free version, the costs associated with accessing its full features may contribute to the digital divide, particularly in underfunded schools. We therefore recommend a practical implementation of the GenT2V AI-generated lesson in a classroom while considering all the limitations and addressing them accordingly.

9. Conclusion and Recommendation

This study assessed the capability of GenT2V AI to enhance teaching for linguistically diverse and culturally varied students. Using VEED, a tool integrated within the OpenAI platform, as MICA, the study explored its potential to address gaps in accessibility and engagement for students with diverse linguistic needs. GenT2V AI demonstrates considerable potential as a tool for creating engaging and accessible lecture content, particularly through its effective captioning features and animated visuals, which enhance student attentiveness and comprehension. These strengths make it a promising resource for classrooms, especially in fostering inclusive and visually appealing presentations. However, critical limitations were observed, particularly in its lack of cultural adaptability, linguistic diversity, and affordability, which restrict its usability in diverse and under-resourced educational contexts. The tool's reliance on Westernized imagery and limited voice-over options fails to align with the cultural and linguistic needs of many global classrooms, diminishing its relevance in settings like Ghana or other underrepresented regions. Additionally, the cost associated with unlocking its full potential through premium features poses significant barriers for educators operating within constrained budgets.

To address these limitations and increase its applicability, GenT2V AI should prioritize improvements in four key areas. First, the incorporation of culturally representative visuals and avatars would ensure content is relatable and inclusive for students from varied backgrounds. Second, expanding the range of language and accent options for voice-overs would cater to global classrooms and multi-lingual students. Third, enhancing accessibility through features such as sign language overlays and customizable audio speeds would improve its utility for students with disabilities or specific learning needs. Finally, offering cost-effective subscription plans tailored for resource-constrained educational settings would make the tool more accessible to a broader audience. While GenT2V AI holds promise as a supplementary teaching tool, addressing these areas would transform it into a universally relevant platform capable of meeting the diverse needs of global educators and students. We recommend an empirical classroom implementation to gather insights from students.

10. Limitation

This study used the free version of GenT2V AI, which came with restricted features. For instance, changing the avatar to the actual image of the instructor was not possible, along with other advanced features. Additionally, while there are many other GenT2V AI tools, we only used VEED, as it was integrated into the OpenAI platform. Although VEED offers advanced features, the findings of this study cannot be generalized to all GenT2V AI tools. Furthermore, many languages were not piloted in this study; we focused only on two languages.

References

1. Wang, X.; Zhang, S.; Yuan, H.; Qing, Z.; Gong, B.; Zhang, Y.; Shen, Y.; Gao, C.; Sang, N. A recipe for scaling up text-to-video generation with text-free videos. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.
2. For Education Statistics [NCES], N.C. English Language Learners in Public Schools, 2021. Available at: <https://nces.ed.gov/programs/coe/indicator/cgf> (Accessed March 1, 2022).
3. Brisk, M.E.; Barnhardt, R.; Herrera, S.; Rochon, R. *Educators' preparation for cultural and linguistic diversity: A call to action*; American Association of Colleges for Teacher Education, 2002. (ERIC No. ED477737).
4. Coady, M.R.; Harper, C.A.; de Jong, E.J. Aiming for equity: Preparing mainstream teachers for inclusion or inclusive classrooms? *TESOL Quarterly* **2016**, *50*, 340–368. doi:10.1002/tesq.252.
5. Cochran-Smith, M.; Lytle, S.L. *Inquiry as stance: Practitioner research for the next generation*; Teachers College Press, 2009.
6. Darling-Hammond, L.; Bransford, J. *Preparing teachers for a changing world: What teachers should learn and be able to do*; Jossey-Bass, 2005.
7. Ponzio, C.M. (Re)Imagining a translingual self: Shifting one monolingual teacher candidate's language lens. *Linguistics and Education* **2020**, *60*, 100866. doi:10.1016/j.linged.2020.100866.
8. Yoon, B.; Pratt, K.L., Eds. *Primary Language Impact on Second Language and Literacy Learning: Linguistically Responsive Strategies for Classroom Teachers*; Lexington Books, 2023.
9. Tarchi, C.; Zaccoletti, S.; Mason, L. Learning from text, video, or subtitles: A comparative analysis. *Computers & Education* **2021**, *160*, 104034. doi:10.1016/j.compedu.2020.104034.
10. Panke, S. Using AI-generated instructor video for multilingual content in MOOCs – An interview with Benedikt Brünner, 2024. Available at: <https://www.aace.org/review>.
11. Dao, X.Q.; Le, N.B.; Nguyen, T.M.T. AI-Powered MOOCs: Video Lecture Generation. *Proceedings of the 2021 3rd International Conference on Image, Video and Signal Processing*; , 2021. doi:10.1145/3459212.3459227.
12. Adetayo, A.J.; Enamudu, A.I.; Lawal, F.M.; Odunewu, A.O. From text to video with AI: The rise and potential of Sora in education and libraries. *Library Hi Tech News* **2024**.
13. Chaugule, B.; Gawade, A.; Mane, P.; Thazhathethil, A.; Kulkarni, S. A Multimodal Journey in Text-to-Image and Video Creation Using AI. *International Research Journal of Innovations in Engineering and Technology* **2024**, *8*, 11.
14. Song, X.; Chen, J.; Zhu, B.; Jiang, Y.G. Text-Driven Video Prediction. *ACM Transactions on Multimedia Computing, Communications, and Applications* **2024**, *20*, Article 296, 15 pages. Available at: <https://doi.org/10.1145/3675171>, doi:10.1145/3675171.
15. Chivileva, I.; Lynch, P.; Ward, T.E.; Smeaton, A.F. A dataset of text prompts, videos and video quality metrics from generative text-to-video AI models. *Data Brief* **2024**, *54*, 110514. doi:10.1016/j.dib.2024.110514.
16. Tan, Z.; Yang, X.; Qin, L.; Li, H. VidGen-1M: A large-scale dataset for text-to-video generation, 2024. doi:10.48550/arXiv.2408.02629.
17. Lucas, T., Ed. *Teacher preparation for linguistically diverse classrooms: A resource for teacher educators*; Routledge, 2010; pp. 75–92.
18. Lucas, T.; Villegas, A.M. Preparing linguistically responsive teachers: Laying the foundation in preservice teacher education. *Theory Into Practice* **2013**, *52*, 98–109. doi:10.1080/00405841.2013.770327.

19. García, S.; Ortiz, A. A framework for culturally and linguistically responsive design of response-to-intervention models. *Multiple Voices for Ethnically Diverse Exceptional Learners* **2008**, *11*, 24–41.
20. Montenegro, E.; Jankowski, N.A. Equity and Assessment: Moving towards Culturally Responsive Assessment, 2017.
21. Câmara, J.N. Funds of knowledge: Towards an asset-based approach to refugee education and family engagement in England. *British Educational Research Journal* **2024**, *50*, 876–904.
22. Scanlan, M. An asset-based approach to linguistic diversity. *Focus on Teacher Education* **2007**.
23. Campos, D. *Educating Latino boys: An asset-based approach*; Corwin Press, 2012.
24. Lindsey, R.B.; Karns, M.; Myatt, K. *Culturally proficient education: An asset-based response to conditions of poverty*; Corwin Press, 2010.
25. Ocumpaugh, J.; Roscoe, R.D.; Baker, R.S.; Hutt, S.; Aguilar, S.J. Toward asset-based instruction and assessment in artificial intelligence in education. *International Journal of Artificial Intelligence in Education* **2024**, pp. 1–40.
26. Flint, A.S.; Jagers, W. You matter here: The impact of asset-based pedagogies on learning. *Theory Into Practice* **2021**, *60*, 254–264.
27. Smith, J. The Ultimate List of STEM Statistics 2024, 2024. Available at: <https://www.codewizardshq.com/stem-statistics/#stem-minorities>.
28. Relyea, J.E.; Amend, S.J. English reading growth in Spanish-speaking bilingual students: Moderating effect of English proficiency on cross-linguistic influence. *Child Development* **2020**, *91*, 1150–1165.
29. Rosheim, K.M.; Tamte, K.G.; Froemming, M.J. Reducing Inequalities Inherent in Literacy Assessment of Multilingual Learners. *Reading Psychology* **2024**. doi:10.1080/02702711.2024.2359922.
30. Hudley, A.H.C.; Mallinson, C. "It's worth our time": a model of culturally and linguistically supportive professional development for K-12 STEM educators. *Cultural Studies of Science Education* **2017**, *12*, 637–660. doi:10.1007/s11422-016-9743-7.
31. Ladson-Billings, G. "Yes, But How Do We Do It?": Practicing Culturally Relevant Pedagogy. In *White Teachers/Diverse Classrooms*; Milner, H.R.; Lomotey, K., Eds.; Routledge, 2023; pp. 33–46.
32. Ladson-Billings, G. Three decades of culturally relevant, responsive, & sustaining pedagogy: What lies ahead? *The Educational Forum* **2021**.
33. Wallace, T.; Brand, B.R. Using critical race theory to analyze science teachers culturally responsive practices. *Cultural Studies of Science Education* **2012**, *7*, 341–374. doi:10.1007/s11422-012-9380-8.
34. Pishghadam, R.; Derakhshan, A.; Zhaleh, K.; Al-Obaydi, L.H. Students' willingness to attend EFL classes with respect to teachers' credibility, stroke, and success: a cross-cultural study of Iranian and Iraqi students' perceptions. *Current Psychology* **2023**, *42*, 4065–4079.
35. Stepp, Z.A.; Brown, J.C. The (lack of) relationship between secondary science teachers' self-efficacy for culturally responsive instruction and their observed practices. *International Journal of Science Education* **2021**, *43*, 1504–1523. doi:10.1080/09500693.2021.1919335.
36. Ghafar, Z.N. ChatGPT: A New Tool to Improve Teaching and Evaluation of Second and Foreign Languages. *International Journal of Applied Research in Sustainability Science* **2023**, *1*, 73–86.
37. Chong, T. Integrating Multimodal Generative AI Technologies in Postgraduate Marketing Education. AS-CILITE Publications, 2024, pp. 37–38. doi:10.14742/apubs.2024.1225.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.