




## Harnessing the power of images in CALL: AI image generation for context specific visual aids in less commonly taught languages

Liang Xu<sup>a</sup>, Elaine Uí Dhonnchadha<sup>b</sup> and Monica Ward<sup>c</sup>

<sup>a</sup>Dublin City University, Dublin, , [liang.xu6@mail.dcu.ie](mailto:liang.xu6@mail.dcu.ie); <sup>b</sup>Trinity College Dublin, Dublin, , [uidhonne@tcd.ie](mailto:uidhonne@tcd.ie); and

<sup>c</sup>Dublin City University, Dublin, , [monica.ward@dcu.ie](mailto:monica.ward@dcu.ie)

How to cite: Xu, L.; Uí Dhonnchadha, E.; Ward, M. (2023). Harnessing the power of images in CALL: AI image generation for context-specific visual aids in less commonly taught languages. In *CALL for all Languages - EUROCALL 2023 Short Papers*. 15-18 August 2023, University of Iceland, Reykjavik. <https://doi.org/10.4995/EuroCALL2023.2023.16950>

---

### Abstract

*This paper explores the application of AI image generation in Computer-Assisted Language Learning (CALL) for Less Commonly Taught Languages (LCTLs). It delves into the potential of text to image generation models in creating context specific visual aids to enhance comprehension and engagement among learners. The integration of AI generated images in a language learning game, CIPHER, is discussed, showcasing the benefits and challenges encountered. Learner feedback indicates positive inclinations towards the AI generated images, but also highlights the need for meticulous selection to address biases and stereotypes. Overall, this approach shows promise in creating culturally relevant CALL resources and improving language learning experiences for learners of LCTLs.*

**Keywords:** CALL, less commonly taught languages, AI image generation, educational games.

---

## 1. Introduction

Images are powerful. With creative ways of harnessing this power, images can be used in education to enhance student engagement. For instance, Callow (2012) presented a pedagogical strategy that leverages students' inherent affinity for images to promote greater participation and involvement in the classroom setting. Research suggests that pictures are more readily identified and retained in memory than words, and visual stimuli are processed at a substantially faster rate compared to textual information (Schroeder et al., 2011). The integration of images and text in Computer-Assisted Language Learning (CALL) can provide numerous advantages, specifically the use of visual imagery can serve as a powerful tool for stimulating learners' interest and engagement, thereby enhancing the overall efficacy of the instructional process.

In the context of CALL for Less Commonly Taught Languages (LCTLs), the incorporation of graphic design elements can be of great benefit. However, due to the relative scarcity of resources available for LCTL CALL development (Ward, 2015) the inclusion of visual design features may be considered a luxury. For instance, in a typical educational game development scenario, the visual elements usually require graphic designers to create imagery based on specific criteria, such as the game's theme, suitability for a certain group of people (e.g. school children), and maintaining cultural relevance. This process is time-consuming and expensive, and many CALL projects focusing on endangered and indigenous languages do not have access to graphic designers due to project

scale and resource limitations. Therefore, we propose AI image generation as a more efficient and cost effective approach to creating context specific imagery in CALL. Image generation tools like Midjourney are able to generate high-quality context specific images based on descriptive prompts.

## **2. Background: Text-to-Image Generation AI for game assets development**

Generative AI, particularly text to image AI has emerged as a powerful tool in game design and the visual arts, enabling the generation of visual content. Text To Image Generation (TTIG) models, such as DALL-E<sup>1</sup>, Midjourney<sup>2</sup>, Stable Diffusion<sup>3</sup>, and Imagen<sup>4</sup>, are self-supervised deep learning models trained on massive datasets. They can generate high quality images based on multi modal commands, including text descriptions and user-provided images. These models enable the manifestation of imaginative concepts, fusion of unrelated objects, and variations of existing images (Ko et al., 2023). The game industry has embraced TTIG models (Vimpari et al., 2023). with the CEO of Unity, a leading game engine company, stating that generative AI will have a transformative impact on gaming (Koetsier, 2023).

Certain models have been trained on open source datasets and made available online for people to access freely (Ko et al., 2023). Moreover, models like Stable Diffusion are open source but require a subscription for access. Vimpari et al. (2023) emphasise the empowering capabilities of TTIG models, enabling users to generate visually appealing content without requiring conventional artistic expertise, which can be financially out of reach for LCTLs with limited resources.

TTIG models find applications across various visual art domains such as graphic design, UI design, webtoons, digital art, and new media art. However, there are limitations - they tend to generate predictable images, lack support for personalisation, and can restrict creativity when solely relying on text prompts (Ko et al., (2023). There are also efficiency considerations and there are concerns about reinforcing stereotypes, biased or harmful content, copyright issues, and artists' authority and compensation (Vimpari, et al., 2023). Despite these limitations and concerns TTIG has firmly arrived in the game industry and visual art field and is expected to impact the creative process from early ideation to the final product's assets (Vimpari, et al., 2023).

## **3. Method: TTIG in CALL for LCTLs**

Images can provide more context for the learning materials delivered through CALL. For example, context assists in reading, and images can be used to enhance context. In our CALL project, Cipher (Ward et al., 2022), a digital game-based language learning application, we employ AI-generated images as visual aids to reinforce reading comprehension and enhance gameplay.



---

<sup>1</sup> <https://openai.com/dall-e-2>

<sup>2</sup> <https://www.midjourney.com>

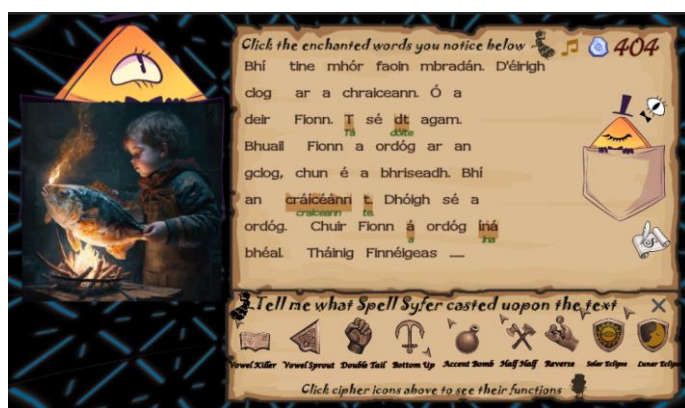
<sup>3</sup> <https://stability.ai/stablediffusion>

<sup>4</sup> <https://imagen.research.google>

Figure 1. Illustrations used in the Cipher game

In Cipher, traditional stories and fairy tales are presented to the learners in pages, with approximately 40 words per page. The images for these pages were generated using an iterative process. Initial prompts for the images were fed into the Midjourney and reviewed. The prompts were then modified to achieve better results. Once an image was deemed suitable, further related images for other parts of the story were generated until all the required images had been generated. Generating images with the same look and feel was a challenge. The illustration (Figure 1 left) is used in the story ‘Hansel and Gretel’ and was generated with the prompt “a witch with her back turned to the fire in her kitchen, with mysterious, transparent background”. The illustration (Figure 1 right) is used in the Irish traditional story “the salmon of knowledge” and was generated with the prompt “mystical boy touching fish roasting on a spit over large fire”. The screenshot displayed in Figure 2 depicts the utilisation of an AI-generated image within the game.

Research and evidence have shown that visual and multi-modal texts are highly effective in captivating and engaging young people across various domains, such as gaming, social media, video creation, and digital text production (Schwienhorst, 2002; Callow, 2012). The utilisation of AI image generation can result in significant improvements to CALL materials, thereby saving time and resources while also reducing costs. The integration of this approach in CALL holds great promise as a powerful educational tool for learners, particularly in the case of LCTL CALL.



Translation:  
 “There was a big fire under the salmon.  
 A blister appeared on its skin.  
 Oh, said Fionn, I've burnt it.  
 He pressed his thumb on the blister to burst it.  
 The skin was hot.  
 He burned his thumb.  
 He put his thumb in his mouth. Finnéigeas ...”

Figure 2. AI generated image in the game along with translation

Furthermore, AI image generation offers the capacity to enhance the game's thematic focus of reconnecting with the spirit of the language by incorporating indigenous and culturally specific elements, thereby fostering a deeper connection with the language's cultural roots (Napier and Whiskeyjack, 2021). In the realm of indigenous mythology and folklore AI image generation has the potential to produce captivating and unique images that can effectively bring stories to life, especially where such imagery may not have been previously available. This is particularly significant for mythological narratives within indigenous cultures, as these stories are often confined to a relatively small global audience. As a result, the circulation of folklore is limited, leading to a scarcity of related imagery. Even when some images are available, they may not be suitable for educational purposes, or may have copyright restrictions.

#### 4. Results and discussion

This section reports on feedback from learners who have used the Cipher game with integrated AI-generated images. Furthermore, we provide guidance advice for other CALL researchers considering the incorporation of AI-generated images into their projects.

#### 4.1. Learner feedback

A user experience investigation, based on a prior study (Xu et al., 2022), focussed on the images incorporated in the game. The participants consisted of individuals aged between 8 and 12 years attending an English medium primary school for boys. A majority of these participants (71%) reported engaging in gaming activities on a daily basis. Following their engagement with the game over two months (once every week, 30 minutes each time) the students were requested to complete a survey. The response categories in the survey were structured using a 5-point Likert scale ranging from 1=very negative to 5=very positive. Responses rated 4 or 5 were categorised as a positive assessment in accordance with the criteria established in Xu et al., (2022).

Of the total 165 responses, 51% expressed a positive inclination towards the story images featured in the game. With respect to the question concerning the extent to which the images contributed to the comprehension of the narrative, 27% of the respondents reported perceiving the story images as facilitative in enhancing their understanding. Additionally, 28% of the responses indicated that the images could potentially assist their comprehension. More than half of the respondents, 53%, expressed that they liked playing the game. A substantial majority of 66% of respondents found the game's approach to learning Irish to be more enjoyable than traditional classroom teaching. Approximately 38% of respondents expressed a positive sentiment toward learning Irish after playing the game and a little over one third of respondents (34%) believed they learned something while playing the game. More detailed information about the survey and its findings can be found in the study (Xu et al., 2023).

#### 4.2. Challenges and suggestions for future work

TTIG AI is currently undergoing rapid development. At the time of writing (July 2023), Google Imagen, a text-to-image diffusion model, is in its beta release phase (Rangwala, 2023). New features which involve extending an image beyond its original borders were implemented in DALL-E in August 2022 (OpenAI, 2022) and Midjourney in June 2023 (Beyer, 2023). Some TTIG models (e.g., DALL-E and Midjourney) have only recently been made available to the public and, at the time of writing, there are almost no regulations currently in place for these tools. TTIG models are part of a larger Artificial Intelligence ecosystem that has been released since 2022 including Generative Artificial Intelligence (GenAI) tools like ChatGPT. Governments and international bodies such as the European Parliament are drafting laws to monitor AI tools. These advancements highlight the dynamic capabilities of image generation while also highlighting certain challenges. The CALL community has continually explored the possibilities offered by new technologies in the language learning process. TTIG AI has the potential to become a powerful tool in the preparation of materials for CALL. It is important to investigate what TTIG AI can provide, particularly as the technology improves, while at all times adhering to national and international regulations.

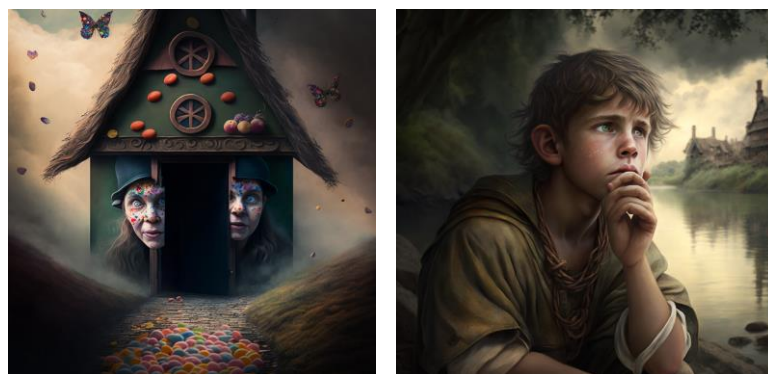


Figure 3. A witch embedded in the front wall (left) and a boy with a strange hand (right)

Throughout the incorporation of AI generated images into the game we faced certain challenges. Some images produced by TTIG models exhibited biases and stereotypes. Therefore, a meticulous selection process is imperative prior to the inclusion of these images in the game. Considerable efforts are required to generate improved images using TTIG models. Achieving specific outcomes often requires numerous attempts involving adjustments to prompts or variations in the generated images. Additionally, it may be necessary to make minor

adjustments using additional editing tools, as some generated images may possess imperfections such as extra fingers on a human subject, see Figure 2 right. Unusual and unwanted ‘blends’, i.e., the blending of objects from the prompts, were a common occurrence. For example, a prompt like "a witch standing in front of her house" could result in an image of a witch embedded in the front wall of a house, see Figure 2 left. Also, the representation of females in the generated images tended to depict stereotypical extremes, portraying women as either youthful and provocative or ancient and repulsive. Any prompts using the word ‘goddess’ were not suitable for a children’s game. It is also difficult to generate a series of pictures for a story which have a common look and style. Nevertheless, with more careful use of prompts and the avoidance of problematic prompt terms very good results are possible.

## 5. Conclusions

This paper explored the use of text-to-image generation models to enhance the development of CALL resources for LCTLs. The positive feedback received from users of the CALL application highlights the potential of AI-generated images in engaging learners and improving comprehension. However, it is crucial to address challenges present in some AI-generated images. There is a need to be aware of the ethical issues with TTIG, particularly the issues of bias and copyright. Continued research and development in this area can further advance the effectiveness and inclusivity of CALL resources, which may ultimately enhance language learning experiences for learners. In summary, although there are challenges in using TTIG models for image generation or ideas drafting, we believe that the benefits make it a worthwhile endeavour.

## Acknowledgements

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No.18/CRT/6224. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

## References

- Beyer, E., (2023). Midjourney 5.2: A guide to the latest tools and features. *NFT Now Podcast* <https://nftnow.com/guides/midjourney-5-2-a-guide-to-the-latest-tools-and-features/>
- Callow, J. (2012). The rules of visual engagement: images as tools for learning. *Screen Education*, (65), 72-79.
- Ko, H. K., Park, G., Jeon, H., Jo, J., Kim, J., & Seo, J. (2023, March). Large-scale text-to-image generation models for visual artists’ creative works. *In Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 919-933). <https://doi.org/10.1145/3581641.3584078>
- Koetsier, J. (2023). Generative AI will completely transform gaming as we know it. *TechFirst with John Koetsier*. <https://johnkoetsier.com/generative-ai-gaming/>
- Napier, K., & Whiskeyjack, L. (2021). wahkotowin: reconnecting to the spirit of nêhiyawêwin (Cree Language). *Engaged Scholar Journal*, 7(1), 1-24. <https://doi.org/10.15402/esj.v7i1.69979>
- OpenAI. (2022). DALL·E: Introducing outpainting. *Open AI*. <https://openai.com/blog/dall-e-introducing-outpainting>
- Rangwala, A., (2023). What is Google Imagen AI? How to use it?. *Open AI Master*. <https://openaimaster.com/google-imagen-ai/>
- Schroeder, S., Richter, T., McElvany, N., Hachfeld, A., Baumert, J., Schnotz, W., Horz, H., & Ullrich, M. (2011). Teachers’ beliefs, instructional behaviors, and students’ engagement in learning from texts

- with instructional pictures. *Learning and Instruction*, 21(3), 403-415.  
<https://doi.org/10.1016/j.learninstruc.2010.06.001>
- Schwienhorst, K. (2002). Why virtual, why environments? Implementing virtual reality concepts in computer-assisted language learning. *Simulation & gaming*, 33(2), 196-209.  
<https://doi.org/10.1177/1046878102332008>
- Vimpari, V., Kultima, A., Hämäläinen, P., & Guckelsberger, C. (2023). "An adapt-or-die type of situation": Perception, adoption, and use of Text-To-Image-Generation AI by Game Industry professionals. *ACM Hum.-Comput. Interact.* 7, CHI PLAY. <https://doi.org/10.1145/3611025>
- Ward, M. (2015). CALL and less commonly taught languages: challenges and opportunities. *Research-publishing.net*. <http://dx.doi.org/10.14705/rpnet.2015.000391>
- Ward, M., Xu, L., & Dhonnchadha, E. U. (2022). Game based language learning for Irish: noticing errors while playing. Intelligent CALL, granular systems and learner data: *short papers from EUROCALL 2022*, 380. <https://doi.org/10.14705/rpnet.2022.61.1488>
- Ward, M., Uí Dhonnchadha, E., McGarry J., and Xu, L. (2023). Co-creating CALL content - does it work? Goldilocks compromise or Cruella chaos? Short papers from EUROCALL 2023.
- Xu, L., Dhonnchadha, E. U., & Ward, M. (2022, June). User experience study of "Cipher: Faoi Gheasa", a digital educational game for language learning and student engagement. *In Proceedings of the 2nd Workshop on Games Systems* (pp. 5-8). <https://doi.org/10.1145/3534085.3534339>