Generative AI in CALL: A September 2024 perspective

Preprint · September 2024

DOI: 10.13140/RG.2.2.28947.49445

3 authors:

Manny Rayner
University of South Australia

Chatgpt C-Lara-Instance

Branislav Bédi
University of Iceland

Generative AI in CALL: A September 2024 perspective[1]

Manny Rayner[1], ChatGPT-4 C-LARA-Instance[1], Branislav Bédi[2]
[1]University of South Australia, Australia
[2]The Árni Magnússon Institute for Icelandic Studies, Iceland

Abstract
Generative AI represents a revolutionary innovation in information technology, impacting many domains including education and computer-assisted language learning (CALL). This short paper explores the development, current applications, and future directions of large language models and large multimodal models in the context of CALL. The perspective is based on the current state of generative artificial intelligence in early September 2024. By using deep learning and reinforcement learning from human feedback, these models demonstrate unprecedented capabilities in generating, understanding, and annotating text and multimedia content. Using examples from recent publications and our own C-LARA platform, we outline the current state of the art, highlighting key functionalities and comparing various generative AI-powered language learning tools, and discuss the challenges in documenting such a rapidly advancing technology. The final section projects likely medium-term developments, arguing that generative AI language tutors will soon achieve near-human proficiency in both text and speech modes. The implications of these advances are profound, promising enhanced language instruction for many, yet raising concerns about digital accessibility and the broader impact of emerging superintelligence.

Keywords: Generative AI (GenAI), Computer-Assisted Language Learning (CALL), ChatGPT, Large Language Models (LLMs), Large Multimodal Models (LMMs), Artificial General Intelligence, superintelligence

[1] Authors in anti-alphabetical order. The three authors made approximately equal contributions to the writing. All content has been reviewed and references added by human authors. A modified version, although written from a slightly different perspective, is also being published in the proceedings of EUROCALL 2024.

1 Introduction

In this paper, our primary concern will be to explore the implications of Large Language Models and Large Multimodal Models for CALL from a September 2024 perspective. It is significant that we feel we need to be so precise about the time point; normally we would have written "2024" or perhaps "mid-2024", but these technologies are evolving at an unprecedented rate, with new developments emerging almost on a weekly basis. Here, we begin with brief definitions of the key concepts, all of which are related to each other:

- **Deep Learning (DL)** is a subfield of machine learning (ML) that uses neural networks with many layers (hence "deep") to analyse various types of data. Deep learning models can automatically learn representations from the input data, making them highly effective for tasks like text, image, and speech processing.
- **Large Language Models (LLMs)** are Deep Learning networks trained on vast amounts of text data. They use deep learning techniques to understand and generate human-like text. These models are pre-trained on diverse datasets and fine-tuned on specific tasks, enabling them to perform a wide range of language-related functions, such as translation, summarization, and conversation in real time. A key component of their training is Reinforcement Learning from Human Feedback (RLHF), which helps refine the model's outputs to better align with human expectations and preferences.
- **Large Multimodal Models (LMMs)** extend the capabilities of LLMs by incorporating multiple data modalities, such as text, images, audio, and video. These models are trained to understand and generate content across these different types of data, allowing for more comprehensive and context aware outputs. LMMs use Deep Learning and RLHF techniques similar to those used for LLMs, but are designed to handle the complexity of multimodal data integration.
- **Reinforcement Learning from Human Feedback (RLHF)** is a training process where models learn from the feedback provided by humans on the quality of their outputs. This iterative process helps fine-tune the model's behaviour to better meet user expectations. It is often not appreciated that, without RLHF, the LLMs are essentially useless.
- **Multimodal Data** involves multiple forms of input, such as text, images, audio, and video. Multimodal models can process and generate content that integrates these different types of data, for example creating images from text descriptions or text descriptions from images.
- **Generative AI (GenAI)** represents AI systems that can create new content, such as text, images, voice, sound or music, rather than just analysing existing data. This capability is powered by LLMs and LMMs, and can generate human-like outputs based on the patterns they have learned during training.
- **Chain of Thought (CoT) Reasoning** is a technique which involves prompting the LLM to "think aloud" when formulating a response. Numerous studies have shown that CoT reasoning can enhance the model's ability to handle complex tasks such as multi-step problem-solving and nuanced language understanding, though the reasons why it works are still somewhat opaque (Feng el al, 2024).

2 Challenges in writing about Generative AI and CALL

As already noted, the field of generative AI is evolving with extraordinary rapidity. The following are key challenges when writing about this technology in the context of CALL:

- **Rapid Technological Advancements and Publication Lag**: Generative AI is advancing so quickly that findings and evaluations can become outdated within months or even weeks. The traditional academic publication process often cannot keep pace with the swift progress in research and development. This lag means that by the time research is published, newer versions of AI models may already have addressed previously identified issues. To take an example from our own work, our December 2023 ALTA paper (ChatGPT et al., 2023) presented evaluation results for language processing in GPT-4. However, GPT-4 Turbo was released shortly before publication, showing strongly improved performance for some languages (Bédi et al, 2024, §5.1.4). Another example is Toby Walsh's book *Faking It* (Walsh, 2023), published in October 2023, which harshly criticised AI capabilities using examples from ChatGPT-3.5 collected in early 2023; by the time of publication, ChatGPT-4 could handle nearly all of the examples correctly (Rayner, 2023). Cases like the above demonstrate the difficulty of capturing the state of the art in this rapidly moving field. This development trajectory suggests that realistic analysis should at the very least assume fast continual improvement along established pathways.
- **Interconnectedness of Generative AI Technology**: Improvements in generative AI capabilities, such as better text understanding and multimodal integration, directly influence CALL applications; unlike the previous generation of CALL platforms, Generative AI based CALL is largely based on generic functionality that can be used in many domains. As a result, discussions must encompass broader AI developments, not just those specific to language learning, to give a proper understanding of the technology's development and impact.
- **Complexity and Interdisciplinarity**: Generative AI for CALL involves different disciplines, including linguistics, computer science, education, cognitive psychology, and ethics. Writing about it requires an interdisciplinary approach to cover the diverse aspects of the technology and its applications effectively. By recognizing the interdisciplinary nature and associated challenges, this paper aims to provide a balanced and forward-looking perspective on the role of generative AI in CALL.

3 Current State of the Art

Generative AI has greatly advanced the capabilities of CALL tools. In this section, we review core issues concerning learner agency and teacher agency and explore some current applications of LLMs and LMMs in CALL, with a particular focus on our own platform, C-LARA.

3.1 Learner Agency and Teacher Agency

Much early work on generative AI and CALL has focussed on the ways in which it interacts with learner agency and teacher agency. On the learner side, generative AI encourages a shift towards pragmatic self-directed learning styles like learner-as-recipient, learner-as-collaborator and learner as-leader (Ouyang and Jiao, 2021), where the learner autonomously or under human teacher guidance selects the AI-powered tools that best suit their learning preferences (Ruiz-Rojas et al., 2023; Boguslawski et al., 2024).

Conversely, the role of human teachers increasingly becomes to help students work effectively in this setting. Teachers need to advise students about generative AI's strengths and limitations and alert them to ethical aspects (Sperling et al., 2024). They need to remind them to use critical thinking, be aware of possible AI-introduced bias, and not become over-reliant on generative AI technologies (Langran et al., 2024, Law, 2024). More powerful CALL tools using generative AI mean that the digital divide between high-tech and low-tech educational settings becomes increasingly significant, placing additional responsibilities on the teacher.

Perhaps most importantly, teacher-student interaction remains a key motivational factor in technology-driven education (Boguslawski, Deer, and Dawson, 2024). There is little evidence so far to suggest that AIs can motivate or be a social support to students as effectively as a human teacher.

3.2 Overview of Current Applications

Various CALL tools and platforms now employ LLMs and LMMs to enhance language learning experiences. Here are some notable examples:

- **Duolingo Max**[2]: Utilizes GPT-4 technology to generate explanations for learners' mistakes, create communicative role-playing tasks, and facilitate real-time spoken and written conversation practice.
- **Memrise**[3]: Employs GPT-3 technology through a chatbot called MemBot, enabling learners to chat, play games, or solve "missions" in real-time communicative tasks.
- **Praktika**[4]: Uses GPT-powered technology for speaking and writing practice with interactive chat characters.
- **Microsoft's AI Immersive Reader**[5]: Supports reading skills by generating translations, highlighting parts of speech, providing illustrations for common words, and splitting words into syllables. Additional features include text-to-speech, line focus, and adjustable reading preferences.
- **Khanmigo**[6]: An online tutoring platform which uses GPT-powered technology to generate lesson plans in different subjects, exercises, and help teachers with administrative tasks. Learners can receive generated guidelines for solving their learning tasks.
- **General-Purpose AI**: Many educational establishments are making AI platforms available to students. To give two examples, the University of Helsinki[7] has for several months now been using a general-purpose Microsoft Copilot assistant with commercial data protection and CurreChat for both the academic staff and the students to assist them with work, research, teaching and study in accordance with the university's principles of AI use (University of Helsinki, 2024). More recently, eight Adelaide schools announced that they

---

[2] https://blog.duolingo.com/duolingo-max/
[3] https://www.memrise.com/blog/introducing-membot
[4] https://praktika.ai
[5] https://azure.microsoft.com/en-us/products/ai-services/ai-immersive-reader
[6] https://www.khanmigo.ai
[7] https://helpdesk.it.helsinki.fi/en/instructions/information-security-and-cloud-services/cloud-services/generative-ai-university

are providing students with free use of EdChat, a ChatGPT based tool with additional safety features.[8]

3.3 Case Study: C-LARA

Since we have full access to the implementation details, the discussion will particularly focus on their own C-LARA[9], a GPT-4o-based open-source platform which enables the creation of multimodal learner texts annotated with glosses, translations, audio, and images. The following key functionalities illustrate its capabilities:

- **Text Generation**: C-LARA can generate a wide variety of engaging learner texts tailored to specific linguistic and pedagogical needs.
- **Text Understanding and Annotation**: Annotation capabilities include segmentation into sentences, addition of lemma tags, first-language (L1) glosses and translations, and construction of “phonetic” versions of texts, where words are decomposed into units associated with phonetic values. Audio annotations for both plain and phonetic texts are created using third-party TTS engines (Bédi et al, 2023, 2024).
- **Multimedia Integration**: The platform can generate images to accompany texts, giving a richer learning experience particularly well suited to beginner/low-intermediate learners. Picture book texts can be created with coherent image sets where the appearance of recurring elements (characters, locations) is kept roughly constant (ChatGPT and Rayner, 2024). This functionality is implemented using GPT-4o’s image generation and image understanding capabilities. The current state of the art already shows a rich landscape of innovative CALL tools powered by generative AI. Next, we consider how one can reasonably expect the field to develop over the next few years.

4 Projected Development

The potential for generative AI in CALL is vast, and ongoing advancements suggest that significant improvements are on the horizon.

4.1 A General AI Language Tutor

The end goal in CALL technology is the development of a general AI language tutor that can engage in unrestricted conversation in both text and speech modes, seamlessly switching between the learner’s first language (L1) and the target language (L2), tracking the learner’s progress, and offering instructional feedback in real time. A tutor of this kind would provide personalized, one-to-one instruction, similar to human tutors. As stressed by Mollick in his influential book *Co-Intelligence* (Mollick, 2024), studies show that personalized tutoring typically offers a two-sigma advantage in learning outcomes, moving an average student into the top 2%.

With the advent of generative AI technology, there now appear to be realistic prospects for actually building such an application. The rationale is straightforward. The key problem previously was that AIs lacked abilities to handle language and common-sense reasoning well enough. This problem has now essentially been solved, at least in text mode. There is nontrivial work left to be done, but the impression is that it can be characterised as good engineering rather than conceptual breakthroughs. We outline what we consider the key issues.

4.2 Current Limitations and Rapid Improvements

[8] https://thepostsa.au/education/2024/08/09/how-ai-is-transforming-sa-schools/
[9] https://www.c-lara.org/

While current AI capabilities, as seen in platforms like Microsoft's Immersive Reader and C-LARA, are impressive, they still have notable limitations:

- **Speech Recognition:** Although text-to-speech technology is well-developed, speech recognition still faces significant hurdles. Understanding of spoken language, especially in noisy environments or with non-native accents, is not yet adequate.
- **Handling Multi-Word Expressions**: Generative AI models often struggle with multi-word expressions (MWEs), which are central to most languages. In experiments with C-LARA, we have for example found that a large proportion of errors in glossing are related to MWEs. Similar behaviour is observed in for example Microsoft's Immersive Reader.
- **Coherent Image Generation**: Creating consistent and contextually accurate images to illustrate texts remains a challenge. In experiments with the August 2024 version of C-LARA, we find that perhaps a third to a half of all generated images are in some way inadequate. Most often this is due to the image generation tool, DALL-E-3, simply not producing an output which matches the prompt.

These limitations are, however, being addressed at a rapid pace. For speech recognition, publicly available information suggests that the problem will be attacked from multiple angles. In particular, tighter integration of speech and text in integrally multimodal models is a principled solution which both OpenAI and Google/Deep Mind are prioritising. OpenAI's Advanced Voice Mode, which supports natural, human-like speech with a latency of around 300ms, was released for alpha testing in Q3 2024 and is expected to become generally available before the end of the year.[10] For CALL, it seems plausible that simply collecting more training data from language learner users and doing more RHLF training in this domain will enable further progress.

An important technique for text processing is Chain of Thought (CoT) reasoning. As of August 2024, we are reengineering the C-LARA platform to incorporate CoT throughout. Initial results on using it for MWE identification are very promising, and similar gains are expected for other language processing tasks; many other groups appear to be using related strategies. More radical solutions are in the pipeline. A high-profile example is OpenAI's "Strawberry"[11], which is widely rumoured to be capable of complex reasoning and will probably be released before the end of 2024. A still more OpenAI advanced platform, "Orion", is supposed to build on Strawberry and include stongly enhanced abilities in language and multimodality, obviously very relevant to CALL applications; it may be released in 2025[12]. Advance information suggests that Strawberry and Orion incorporate techniques which allow the AIs to generate their own training material to fine-tune CoT reasoning, probably using techniques related to the Self-Taught Reasoner (StaR) algorithm (Zelikman et al, 2022).

An interesting development is the rapid progress in text-to-music generation, as exemplified in platforms like Suno[13] and Udio[14]. This technology is good enough that an Udio-generated song, in

10 https://www.theverge.com/2024/8/15/24220378/openai-advanced-voice-mode-uncanny-valley
11 https://www.reuters.com/technology/artificial-intelligence/openai-working-new-reasoning-technology-under-code-name-strawberry-2024-07-12/
12 https://decrypt.co/247769/openai-strawberry-orion-gpt-next
13 https://suno.com/
14 https://www.udio.com/

August 2024, made the top 50 on the German music list.[15] There is, again, obvious applicability to CALL, and we are one of many groups currently working to integrate it into our platform.

4.3 Timeline for Advancements

Following straightforward trendline analysis, such as that presented in (Aschenbrenner, 2024), it is reasonable to expect Artificial General Intelligence (AGI) before the end of the decade. Based on such projections, a general AI language tutor capable of near-human proficiency in both text and speech modes may well be realized within 2–4 years, i.e. 2026–2028.

5 Future Functionalities and Superintelligence

As generative AI continues to evolve, one can expect the emergence of new functionalities that will further enhance CALL tools. The following are particularly interesting:

- **Video Generation**: While still in its infancy, video generation technology is expected to advance over the next few years, providing richer multimedia material for language teachers and learning content for language learners.
- **Embodied Intelligence**: The development of generative AI-powered robots and virtual avatars could offer even more immersive language learning environments, simulating unlimited real-life interactions.

Looking further ahead, the advent of superintelligence has been predicted by many people (Bostrom, 2014). Aschenbrenner (2024), projects that it will arrive a year or two after the AGI milestone has been reached, hence around 2029–2030. Such claims seem less speculative when we recall that superhuman performance is already available in very challenging domains such as chess and Go (Silver et al, 2018) and protein folding (Jumper et al, 2021). Analysis of AIs like Deep Mind's AlphaZero (chess and Go) abundantly confirms that they do not just *achieve results* better than any humans, but also *understand* these domains far better (Sadler and Regan, 2019). If general superintelligence does indeed arise, it will bring about profound changes not just in CALL but across all aspects of society, which are impossible to anticipate even in broad outline.

In the medium-term, it is reasonable, as outlined above under "Learner agency and teacher agency", to project that powerful CALL tools using generative AI will make learners increasingly autonomous. The teacher's task will more and more become to help the learners use these tools appropriately. Their ability to motivate, engage, and support students, combined with their expertise in guiding ethical and critical use of AI tools, is needed to ensure that the integration of AI in education leads not just to technical successes but to positive outcomes at the human level.

References

Aschenbrenner, L. (2024). *Situational Awareness*. https://situational-awareness.ai/.

Bédi, B., ChatGPT-4-Instance, Chiera, B., Chua, C., Ní Chiaraín, N., Rayner, B., Simonsen, A., Zviel-Girshin. (2023). ChatGPT-Based Learning And Reading Assistant: Initial Report. ResearchGate. Preprint. http://dx.doi.org/10.13140/RG.2.2.12715.82721.

Bédi, B., ChatGPT-4-Instance, Chiera, B., Chua, C., Cucchiarini, C., Dotte, A-L., Geneix-Rabault, S., Maizonniaux, Ch., Marginean, C., Ní Chiaraín, N., Parry-Mills, L., Raheb, Ch., Rayner, B., Simonsen, A., Viorica, M., L., Wacalie, F., Welby., P., Xiang, Z., Zviel-Girshin. (2024). ChatGPT-

[15] https://www.watson.ch/leben/digital/405756915-verknallt-in-einen-talahon-ki-song-sorgt-fuer-stirnrunzeln

Based Learning And Reading Assistant (C-LARA): Second Report. ResearchGate. Preprint. http://dx.doi.org/10.13140/RG.2.2.28235.40480.

Boguslawski, S., Deer, R. and Dawson, M. G. (2024). Programming education and learner motivation in the age of generative AI: student and educator perspectives. *Information and Learning Sciences*. https://doi.org/10.1108/ILS-10-2023-0163.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

ChatGPT C-LARA-Instance, Chiera, B., Chua, C., Raheb, Ch., Rayner, M., Simonsen, A., Xiang, Z., and Zviel-Girshin, R. (2023). Using C-LARA to evaluate GPT-4's multilingual processing. In *Proceedings of the 21st Annual Workshop of the Australasian Language Technology Association*, pages 20–29, Melbourne, Australia. Association for Computational Linguistics. https://aclanthology.org/2023.alta-1.3.pdf.

ChatGPT, C. and Rayner, M. (2024). Making Picture Book Texts with C-LARA (interim report). ResearchGate. Preprint. http://dx.doi.org/10.13140/RG.2.2.22001.70248.

Fengchun, M. and Wayne, H. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535.

Feng, G., Zhang, B., Gu, Y., Ye, H., He, D., Wang. L. (2024). Towards revealing the mystery behind chain of thought: a theoretical perspective. *Advances in Neural Information Processing Systems 36*. https://doi.org/10.48550/arXiv.2305.15408.

Jumper, J., Evans, R., Pritzel, A. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Langran, E., Searson, M. and Trumble, J. (2024). Transforming Teacher Education in the Age of Generative AI. In Searson, M. Langran, E., Trumble J. (Eds.), *Exploring new horizons: Generative artificial intelligence and teacher education*. Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/p/223928/.

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open, 6.* Elsevier. https://doi.org/10.1016/j.caeo.2024.100174.

Mollick, E.(2024). *Co-Intelligence*. Random House UK.

Ouyang, F., and Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence 2*. https://doi.org/10.1016/j.caeai.2021.100020.

Rayner, M. (2023). Review of *Faking It: Artificial Intelligence in a Human World*. https://www.goodreads.com/review/show/5915548332.

Ruiz-Rojas, L.I.; Acosta-Vargas, P.; De-Moreta-Llovet, J.; Gonzalez-Rodriguez, M. (2023). Empowering Education with Generative Artificial Intelligence Tools: Approach with an Instructional Design Matrix. *Sustainability*, *15*(11524). https://doi.org/10.3390/su151511524.

Sadler, M. and Regan, N. (2019). *Game Changer: AlphaZero's Groundbreaking Chess Strategies and the Promise of AI*. New in Chess.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L. Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., and Hassabis, D., (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science, 362*(6419):1140–1144. https://doi.org/10.1126/science.aar6404.

Walsh, T. (2023). *Faking It: Artificial Intelligence in a Human World*. Black Inc.

Zelikman E., Wu Y., Mu J., Goodman N. (2022) Star: Bootstrapping reasoning with reasoning. Advances in Neural Information Processing Systems.