



Technical recommendations for Artificial Intelligence and Augmented Reality tools to support language learning

A guide for developers, instructional designers and language teachers



Authors: Fridolin Wild, Mikhail Fominykh, Andreas Kißmehl,
Tormod Aagaard, Antigoni Parmaxi, Eirini Christou, Anke Berns,
Lina Adinolfi, Skevi Vassiliou, and Angeliki Voreopoulou

The Augmented Reality Instructional Design for Language Learning – ARIDLL project has received funding from the European Union's Erasmus Plus programme, grant agreement 2022-1-NO01-KA220-HED-000088034.



Document

Technical recommendations for Artificial Intelligence and Augmented Reality tools to support language learning: A guide for developers, instructional designers and language teachers

Authors

Editor: Fridolin Wild

Contributors: Fridolin Wild, Mikhail Fominykh, Andreas Kißmehl, Tormod Aagaard, Antigoni Parmaxi, Eirini Christou, Anke Berns, Lina Adinolfi, Skevi Vassiliou, and Angeliki Voreopoulou

Acknowledgement

The Augmented Reality Instructional Design for Language Learning – ARIDLL project has received funding from the European Union’s Erasmus Plus programme, grant agreement [2022-1-NO01-KA220-HED-000088034](https://ec.europa.eu/erasmus-plus/en/2022-1-NO01-KA220-HED-000088034).

Disclaimer

The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.



Co-funded by the
Erasmus+ Programme
of the European Union

License

This document is published by the ARIDLL consortium with a [CC-BY-SA-4.0](https://creativecommons.org/licenses/by-sa/4.0/) license.



Cite as

Fridolin Wild, Mikhail Fominykh, Andreas Kißmehl, Tormod Aagaard, Antigoni Parmaxi, Eirini Christou, Anke Berns, Lina Adinolfi, Skevi Vassiliou, and Angeliki Voreopoulou (2025) Technical recommendations for Artificial Intelligence and Augmented Reality tools to support language learning: A guide for developers, instructional designers and language teachers. ARIDLL Consortium. <https://aridll.eu/>

Table of contents

| | |
|---|-----------|
| Table of contents | 3 |
| Executive summary | 5 |
| Key terminology | 6 |
| 1. Introduction | 7 |
| 1.1 Who this guide is for | 10 |
| 1.2 The methodology behind this guide | 10 |
| 1.3 Underlying design principles | 11 |
| 1.4 MirageXR | 14 |
| 1.5 How to read this guide | 15 |
| 2. Prompt patterns and a pattern language | 16 |
| 2.1 The pattern format | 17 |
| 2.1.1. Context | 18 |
| 2.1.2. Knowledgebase | 19 |
| 2.1.3. The user prompt (also known as the message) | 19 |
| 2.2 A prompt engineering template | 20 |
| 2.3 Pattern language | 21 |
| 3. Technical considerations | 22 |
| 3.1 Chain of Thought | 22 |
| 3.2 Few Shot (FS), One-shot (1S), Zero-shot (0S) | 23 |
| 3.3 Multitask learning | 23 |
| 3.4 Retrieval-Augmented Generation | 24 |
| 3.5 Memory | 25 |
| 3.6 Fine-tuning | 25 |
| 3.7 Restrictions arising from web communication | 26 |
| 4. Language related patterns | 27 |
| 4.1 Prompt language | 27 |
| 4.2 Language proficiency levels (from A1 to C2) | 28 |
| 4.3 Focusing on specific language elements | 30 |
| 4.3.1 Phonetics | 31 |
| 4.3.2 Phonology | 31 |
| 4.3.3 Orthography | 33 |
| 4.3.4 Morphology | 33 |
| 4.3.5 Syntax | 34 |
| 4.3.6 Semantics | 34 |
| 4.3.7 Pragmatics | 35 |
| 4.3.8 Co-speech gestures | 35 |
| 4.4 Other modalities and multimodal inputs | 36 |
| 5. Pedagogical patterns for language learning | 38 |
| 5.1 Meta Prompts and options (to provide more learner autonomy) | 38 |
| 5.2 Guiding by examples | 39 |
| 5.3 Gamification | 40 |

| | |
|---|-----------|
| 5.4 Feedback with recasting, corrective repetition, and explanations..... | 40 |
| 5.5 Roles, personas, and where to find them..... | 44 |
| 5.6 Keeping the conversation going..... | 45 |
| 5.7 Ending the conversation..... | 45 |
| 6. Patterns for situated task practice..... | 46 |
| 6.1 Ice-breakers..... | 46 |
| 6.2 Small talk..... | 46 |
| 6.3 Information gathering..... | 46 |
| 6.4 Employability..... | 47 |
| 6.5 Public Speaking practice..... | 47 |
| 6.6 Negotiation skills..... | 47 |
| 6.7 Travelling and real-life situations..... | 48 |
| 6.8 Guessing games..... | 48 |
| 6.9 Vocabulary training games..... | 48 |
| 7. Conclusion..... | 49 |
| 7.1 Sense Reason Act..... | 49 |
| 7.2 Model Context Protocol..... | 49 |
| 7.3 Middleware..... | 49 |
| 7.4 Automated help and guidance..... | 49 |
| 7.5 Wider range of languages..... | 50 |
| 7.6 Prompt language..... | 50 |
| 7.7 Bias..... | 50 |
| 7.8 Privacy..... | 50 |
| 7.9 Theory-aware agents..... | 50 |
| 7.10 Multimodal learning..... | 51 |
| 7.11 Piloting: larger, longer, and more languages..... | 51 |
| 7.12 Summary..... | 51 |

Executive summary

This guide provides technical recommendations for using Artificial Intelligence (AI) and Augmented Reality (AR) tools to support instructed language learning. It is primarily aimed at application developers, instructional designers and language teachers who utilise AI technologies and designed experiences, though parts may also benefit learners interested in enhancing their media competence.

With this guide we reflect on a series of experiments, pilot studies and expert workshops, to gather the knowledge we have built in using and teaching these technologies. We have explored the longitudinal use of AI+AR in our Erasmus+ funded project for almost three years (at the time of writing), and this guide provides a summary of our insights on both what works and what to pay attention to.

The guide is based on principles grounded in psychological research, user experience, and pedagogical good practice. It prioritises user experience, drawing on frameworks such as the Motivation, Engagement and Thriving in User Experience (METUX) model, which focuses on psychological needs such as Autonomy, Competence, and Relatedness. The CLEAR Framework (Concise, Logical, Explicit, Adaptive, Reflective) is also adopted to improve interactions with AI language models.

The guide is divided into two key parts:

Part A provides an overview of the role and benefits of AI-powered AR. AR overlays computer-generated images onto the real world, while AI enables computer systems to perform tasks requiring human intelligence, such as reasoning and problem-solving. The fusion of AI into AR applications creates a powerful and personalised educational tool, capable of analysing a learner's proficiency and tailoring learning experiences to their needs. This can include AI-powered virtual humans providing personalised feedback within immersive, real-world scenarios for conversational practice.

Part B in turn focuses on prompt engineering. It introduces prompt patterns and a pattern language for designing AI+AR prompts, modifying traditional design pattern formats to include key elements such as context, knowledgebase, and user prompt. These are crucial for configuring AI+AR-enabled virtual humans with specific conversational traits and behaviours. This part covers good practice in writing prompts, technical considerations like Chain of Thought (CoT), Few-Shot learning (FS), Retrieval-Augmented Generation (RAG), and Memory. It also details language-related patterns and pedagogical patterns for language learning, including feedback strategies, roles/personas, and conversation management. It concludes with patterns for situated task practice in various real-life scenarios.

Key terminology

Augmented Reality (AR): Refers to overlaying real world views with interactive 3D digital content with the user freely choosing their viewpoint.

Artificial Intelligence (AI): AI is a broad field of computer science that is concerned with creating machines that can perform tasks that typically require human intelligence.

Conversational agent: Often used interchangeably with terms like chatbot, virtual assistant, or dialogue system, this is a computer program designed to interact with humans using natural language. The primary goal of a conversational agent is to simulate a human-like conversation, whether through text, voice or a combination of both.

Large language model (LLM): This is a type of AI model trained on vast amounts of textual (and often other, multimodal) data to understand, interpret and generate human language.

Prompt: A prompt is the input or instruction given to an AI model to obtain a specific output. It can be both (1) a question that requires an answer or (2) information to inform possible answers (like a command or set of examples).

Prompt engineering: Prompt engineering is the practice of carefully crafting and optimizing prompts to guide an AI model toward generating the desired output.

RAG (Retrieval-Augmented Generation): RAG is a technique used to improve the accuracy and relevance of AI models, particularly LLMs. It works by giving the AI access to an external knowledgebase (like a private database or a set of documents) before it generates a response. This allows the AI to 'retrieve' up-to-date and specific information from this external source and incorporate it into its answer.

Virtual human: This is a software-generated, three-dimensional character or human being that is designed to look, move and behave similarly to a real person. These digital entities are created using a combination of computer graphics, AI and other technologies to enable realistic interaction and communication.

1. Introduction

Augmented Reality (AR) is a technology that overlays computer-generated images onto the real world, enhancing the user's perception of their environment¹. AR blends digital elements with the user's existing surroundings, often through a mobile device (smartphone, tablet) or specialized glasses. AR serves so-called 'reality-embedded learning', serving competence development through conversational information exchange enriched and enacted with experience (see Figure 1).

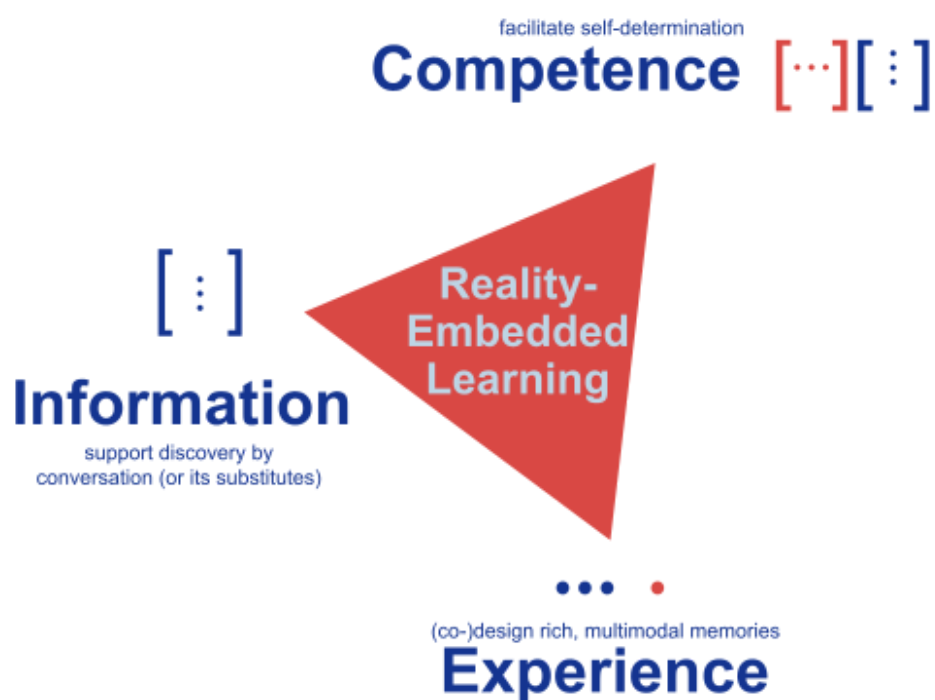


Fig.1. Reality-embedded learning.

AR experiences can be significantly enhanced through the integration of Artificial Intelligence (AI), which provides the intelligent processing capabilities necessary for real-time object recognition, contextual understanding and adaptive content delivery.

AI refers to the ability of computer systems to perform tasks that typically require human intelligence, such as reasoning, problem-solving and decision-making. Key AI-powered tools for instructed language acquisition include speech recognition software, intelligent virtual tutors, machine translation, writing assistants, and educational games². Zhu et al. (2024)³ noted that AI use for language education prevalently used automated writing evaluation, bots, machine translation, automatic speech recognition and intelligent systems.

The use of AI and chatbots in language learning has attracted considerable scholarly interest, resulting in a substantial body of literature discussing its opportunities in enhancing the teaching

¹ Dargan, S., Bansal, S., Kumar, M., Mittal, A., & Kumar, K. (2023). Augmented reality: A comprehensive review. Archives of Computational Methods in Engineering, 30(2). <http://doi.org/10.1007/s11831-022-09831-7>.

² Alshumaimeri, Y. A.; Alshememry, A. K. (2024): The Extent of AI Applications in EFL Learning and Teaching, in: IEEE Transactions on Learning Technologies, vol. 17. <https://doi.org/10.1109/TLT.2023.3322128>

³ Zhu, M.; Wang, C. (2024): A Systematic Review of Artificial Intelligence in Language Education from 2013 to 2023: Current Status and Future Implications. In: SSRN. <http://dx.doi.org/10.2139/ssrn.4684304>

and learning process (Frances & Zimotti, 2023⁴; Godwin-Jones et al, 2024⁵; Roose, 2023⁶; Tseng & Warschauer, 2023⁷, Weissmann, 2023⁸, Greene, 2022⁹). In the context of instructed language acquisition, such opportunities can be found in the creation of student-centred tasks for practising writing and speaking skills, enabling students to extend their practice of the target language beyond the classroom (Şahin Kızıl et al., 2025¹⁰). Chatbots can be used as conversational partners that are available at any time, including outside classroom and tutoring hours (Godwin-Jones, 2024¹¹). This enables students to practise independently, according to their language level and learning needs. Prompts can be created by either teachers or the students themselves, based on the topics, communicative modes and elements they wish to practise, while complementing any existing syllabus or coursebook. AI algorithms can analyze a learner's language proficiency, identify areas of improvement and tailor the experience to precise needs. The integration of AI into AR applications in language learning thus has the potential to create a powerful, personalized educational tool (Geroimenko, 2023)¹².

More specifically, AI-powered virtual humans (see Figure 2) can provide multi-modal feedback on different aspects of communication embedded in immersive, real-world scenarios that practise conversational skills. By combining the immersive and interactive nature of AR with the adaptive capabilities of AI, language learning can be both contextually relevant and progressively challenging, and potentially more engaging, enjoyable and effective (Iqbal et al., 2022)¹³.

Elements of language (see Parmaxi et al., 2025, Instructional Design Guide¹⁴) refer to: vocabulary, pronunciation, spelling, morphology, grammar, phraseology, meaning, and rhetoric and culture. In Section 4.3, we explain these in more detail.

⁴ Frances, C., & Zimotti, G. (2023). Robots versus humans: Does ChatGPT pose a challenge to second language writing? The FLTMAG. <https://fltmag.com/chatgpt-second-language-writing>

⁵ Godwin-Jones, R., O'Neill, E., & Ranalli, J. (2024). Integrating AI tools into instructed second language acquisition. In Exploring artificial intelligence in applied linguistics, Iowa State University Digital Press. <https://doi.org/10.31274/isudp.2024.154.02>

⁶ Roose, K. (2023). [Don't ban ChatGPT in schools. Teach with it.](#) The New York Times.

⁷ Tseng, W. T., & Warschauer, M. (2023). AI-writing tools in education: If you can't beat them, join them. Journal of China Computer-Assisted Language Learning, 3(2). <https://doi.org/10.1515/jccall-2023-0008>

⁸ Weissman, J. (2023). [ChatGPT is a plague upon education.](#) Inside Higher Ed.

⁹ Greene, P. (2022). [No. ChatGPT is not the end of high school English. But here's the useful tool it offers teachers.](#) Forbes.

¹⁰ Şahin Kızıl, A., Klimova, B., Pikhart, M., & Parmaxi, A. (2025). A systematic review of the recent research on the usefulness of chatbots for language education. Journal of Computer Assisted Learning, 41(2). <https://doi.org/10.1111/jcal.70001>.

¹¹ Godwin-Jones, R., 2024. Distributed agency in second language learning and teaching through generative AI. Language Learning & Technology, 28(2), 5–30 <https://doi.org/10.64152/10125/73570>.

¹² Geroimenko, V. (Ed.). (2023). Augmented reality and artificial intelligence: The fusion of advanced technologies. Springer Nature. <https://doi.org/10.1007/978-3-031-27166-3>.

¹³ Iqbal, M. Z., Mangina, E., & Campbell, A. G. (2022). Current challenges and future research directions in augmented reality for education. Multimodal Technologies and Interaction, 6(9). <https://doi.org/10.3390/mti6090075>.

¹⁴ Antigoni Parmaxi, Anke Berns, Lina Adinolfi, Alice Gruber, Mikhail Fominykh, Angeliki Voreopoulou, Fridolin Wild, Paraskevi Vassiliou, Eirini Christou, Concepción Valero-Franco, Tormod Aagaard, Stella Hadjistassou, and Pedro Jesús Molina Muñoz (2025) Instructional design guide for Augmented Reality in language learning. ARIDLL consortium. <https://aridll.eu/>

In the era of AI-powered AR, there is increasing recognition of the value to teachers of mastering sophisticated prompt engineering skills to harness intelligent systems that can generate contextually relevant immersive learning experiences tailored to diverse student needs and curricular objectives. Isemonger (2023)¹⁵ even makes the case to reposition educators as specialized ‘educational prompt engineers’, skilled at guiding students in effectively interacting with generative AI to enhance learning, not hinder. Teachers can leverage these AI tools to offer extensive conversational practice, personalized corrective feedback, and linguistic variation critical for language acquisition. Isemonger (2023) highlights that successful integration requires precise prompting skills, i.e., making explicit the instructions that typically are implicit in human interaction (i.e. cues that rely heavily on shared context, pragmatics and social conventions).

Kosmina et al. (2025) argue that using AI to enhance human cognition is beneficial, while relying on AI as a substitute for cognitive effort leads to the accumulation of “cognitive debt”, severely limiting neural connectivity and impairing memory to a level where test participants cannot remember what they wrote ten minutes ago¹⁶.

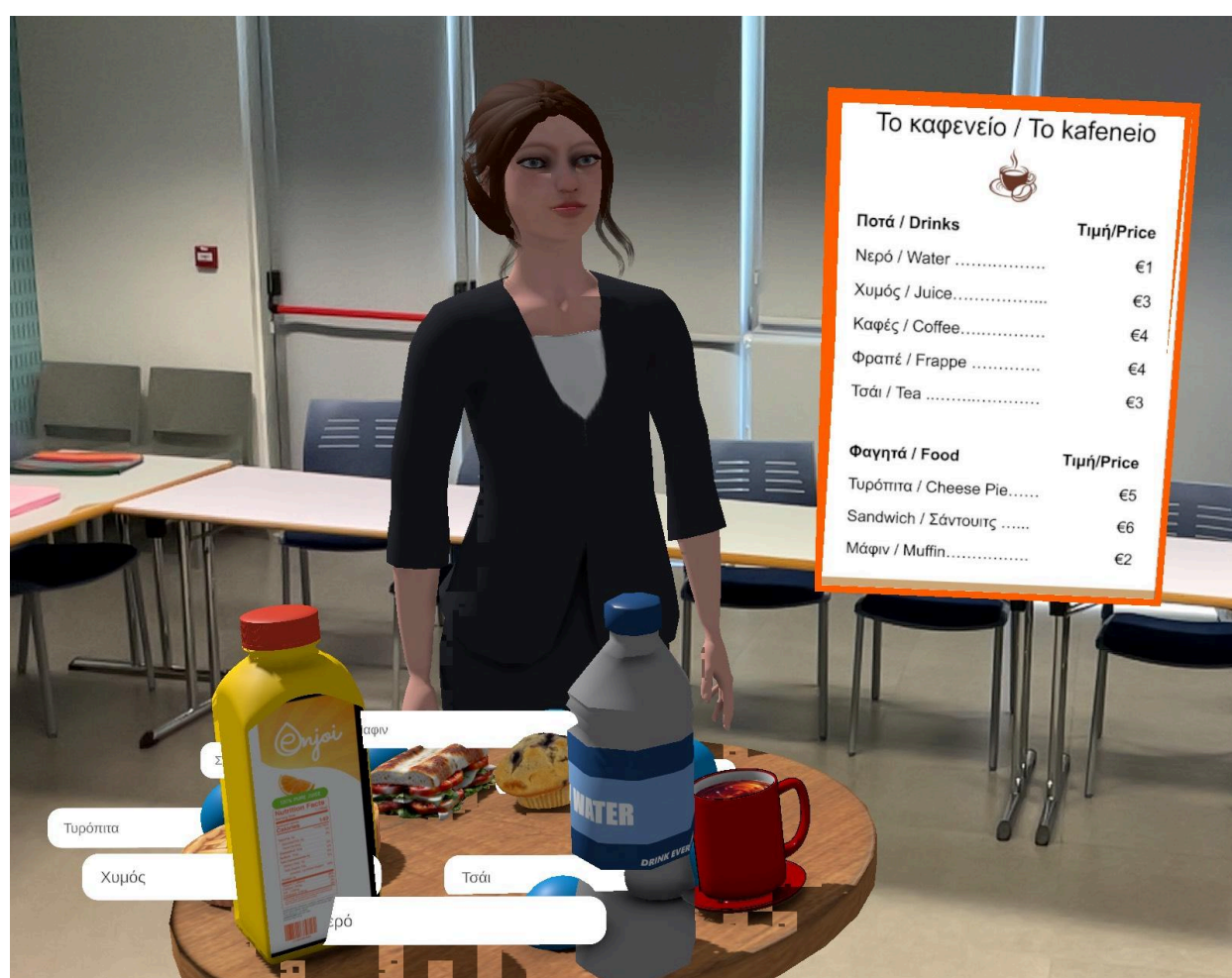


Fig.2. AI+AR in action: A virtual human in a Greek cafe (image: Cyprus University of Technology)

¹⁵ Isemonger, I. (2023) Generative Language Models in Education: Foreign Language Learning and the Teacher as Prompt Engineer, In: TEFL Praxis Journal, Vol. 2. <https://doi.org/10.5281/zenodo.10402411>.

¹⁶ Nataliya Kosmyna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, Pattie Maes (2025): Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task, <https://arxiv.org/abs/2506.08872>.

1.1 Who this guide is for

The intended audience of this guide embraces application developers, instructional designers and language teachers who are interested in using AI technologies and designed experiences. Parts of the guide will also be of interest to learners who want to benefit from the technology and enhance their media competence (AI application skills). This diverse group includes school students, students in Higher Education and adult learners.

1.2 The methodology behind this guide

This guide is the product of a participatory, layered approach involving AI+AR experts and language teachers who reported back from a series of events and pilot trials. These included initial ideation workshops where participants established common ground, identified shared objectives and aligned their diverse perspectives on how AI-powered AR technology could be integrated in the language classroom. Subsequent trial implementations took place across different countries within the ARIDLL consortium in order to test different aspects of the evolving AI+AR language learning methodology and applications. These pilots involved diverse educational settings, different learner age groups, proficiency levels and learning contexts. Each pilot focused on specific functionalities: for example, one examined AI-powered feedback, another examined vocabulary acquisition etc. Throughout these pilots, the technical team focused on the design of the AI+AR system (MirageXR, see section 1.4), while the pedagogical team translated technical possibilities into educationally sound practices, defining learning objectives, and providing appropriate content. The debriefing mechanism ensured continuous knowledge transfer. This participatory, layered approach reflects best practice in educational technology research, where successful innovation is not simply a question of technical advancement but requires careful consideration of learning design. The approach adopted is described in detail below:

First, targeting language teachers in particular, two rounds of content authoring and experimentation workshops were delivered in the consortium members' respective countries. The aim of these workshops was to train instructors on the use of the MirageXR application and the development of AR language learning activities to implement with their students as part of their language courses. The resulting AR+AI scenario activities were piloted subsequently in different local institutions. Students were observed while interacting with the activities and necessary guidance was provided. Feedback received both from students and instructors was used to refine the AR application, the AI prompts used and the patterns extracted from them into this guide.

This was accompanied, secondly, with expert workshops, comprising all consortium members, where the content produced, problems, solutions and strategies were further discussed, resulting in several comprehensive prompt collections.

Third, we initiated a regular series of prompt engineering workshops for consortium members, over the course of which we started extracting patterns and working towards the development of this guide (Fig. 3).

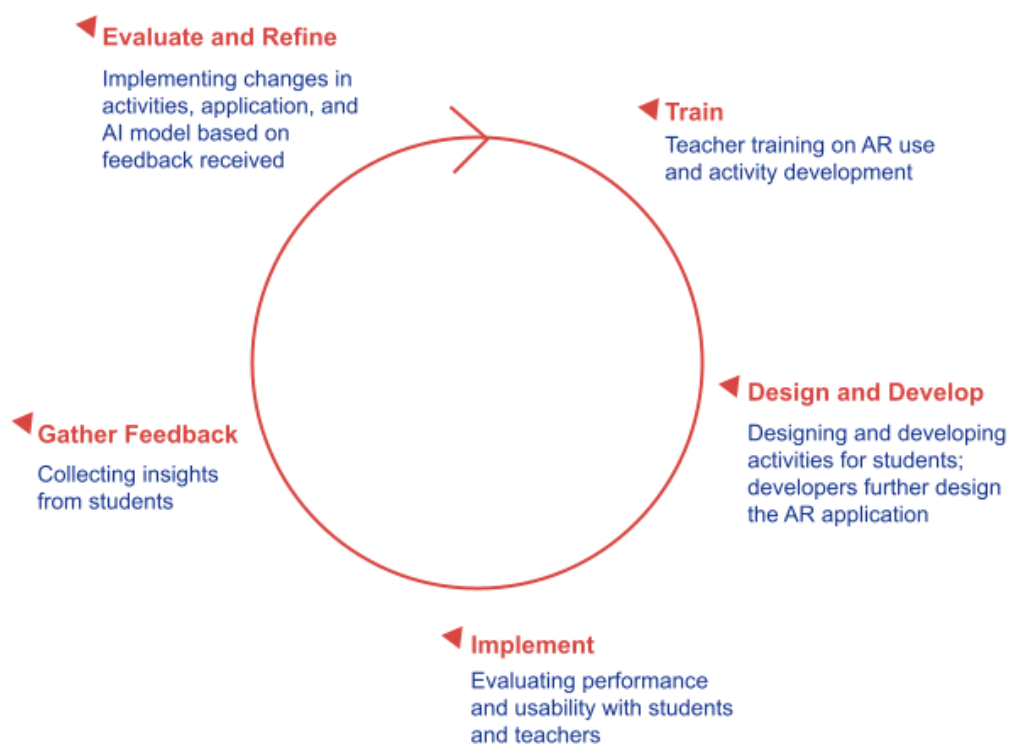


Fig.3. Methodology underlying this guide (own figure).

1.3 Underlying design principles

Our underlying design principles serve as a crucial bridge between established language learning theories and emerging AI+AR technologies. To prevent technological novelty from dictating our educational approach, our design principles are firmly grounded in pedagogical best practices, psychological research, and user experience (UX).

The conceptualisation and design of AR and AI-powered applications should prioritise user experience over mere hardware capabilities (Rauschnabel et al., 2022)¹⁷. This is corroborated by Parmaxi and Demetriou (2020)¹⁸, who highlighted the importance of aligning AR affordances with 21st-century skills and addressing areas like instructional design and user experience.

User experience is a multidimensional, complex construct influenced by the user's internal state (e.g., needs, moods, expectations), the characteristics of the technology (e.g., complexity, functionality, usability), and the interaction environment (Rauschnabel et al., 2022)¹⁹.

¹⁷ Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H., & Alt, F. (2022). What is XR? Towards a framework for augmented and virtual reality. *Computers in human behavior*, 133. <https://doi.org/10.1016/j.chb.2022.107289>.

¹⁸ Parmaxi, A., & Demetriou, A. A. (2020). Augmented reality in language learning: A state-of-the-art review of 2014–2019. *Journal of Computer Assisted Learning*, 36(6). <https://doi.org/10.1111/jcal.12486>.

¹⁹ Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H., & Alt, F. (2022). What is XR? Towards a framework for augmented and virtual reality. *Computers in human behavior*, 133. <https://doi.org/10.1016/j.chb.2022.107288>.

The **Motivation, Engagement and Thriving in User Experience (METUX)** model offers a helpful framework for understanding how the design of technology in general can facilitate satisfaction of psychological needs across various spheres of user experience (Peters et al., 2018)²⁰. The framework is grounded in psychological research, specifically the Self-Determination Theory, which identifies three basic psychological needs crucial for self-motivation and wellbeing, namely: Autonomy (feeling of agency), Competence (feeling able and effective), and Relatedness (a sense of belonging). The model (see Figure 4) posits that these needs must be considered across separable spheres of experience, i.e., interface, task, behavior and life, to ensure a broad view of wellbeing and to prevent the creation of designs that might satisfy needs at one level while undermining them at another, potentially leading to issues such as addiction.

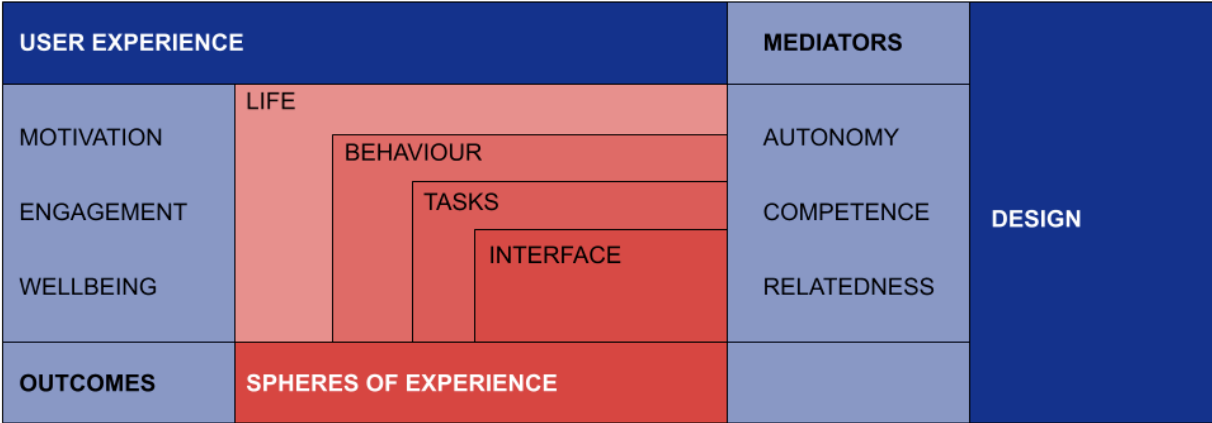


Fig.4. METUX model (redrawn and simplified from Peters et al., 2018)

METUX provides guidance for AI+AR language learning researchers and practitioners to design technologies that genuinely increase motivation, engagement and user wellbeing by focusing on their underlying fundamental psychological needs. We recommend to deploy METUX in the design of user experience, as also illustrated in Figure 5.

We also apply the CLEAR Framework²¹ to ensure AI interaction is structured and effective. CLEAR offers another, complementary method to METUX, more focused on improving interaction with the AI large language models like ChatGPT. CLEAR postulates five core principles for prompt development:

- **Concise:** Keep prompts short and clear, avoiding unnecessary details to generate precise, relevant responses.
- **Logical:** Structure prompts logically to ensure that the AI understands the relationships and context clearly.
- **Explicit:** Clearly define the expected format (i.e. length, formality, style), scope and content to avoid vague or irrelevant responses.
- **Adaptive:** Adjust prompts flexibly, experimenting with phrasing or settings to balance creativity and accuracy.

²⁰ Peters, D., Calvo, R. A., & Ryan, R. M. (2018). Designing for motivation, engagement and wellbeing in digital experience. *Frontiers in psychology*, 9, <https://doi.org/10.3389/fpsyg.2018.00797>.

²¹ Lo, L. (2023): The CLEAR path: A framework for enhancing information literacy through prompt engineering, In: *The Journal of Academic Librarianship* 49 (2023) <https://doi.org/10.1016/j.acalib.2023.102720>.

- **Reflective:** Continuously evaluate AI responses to refine future prompts for better outcomes.

In this guide, we maintain the user at the centre, as suggested by the Self-Determination Theory design needs, to enhance the user's sense of agency, willingness to engage, and the alignment of the available activities with their goals and values. This includes providing choices, options, and personalisation features in the prompt examples and the patterns identified. We demonstrate how our patterns facilitate motivation, engagement, and wellbeing, and how they can provide more flexibility in the mediators (autonomy, competence, relatedness), to acknowledge the significance of this user experience perspective.

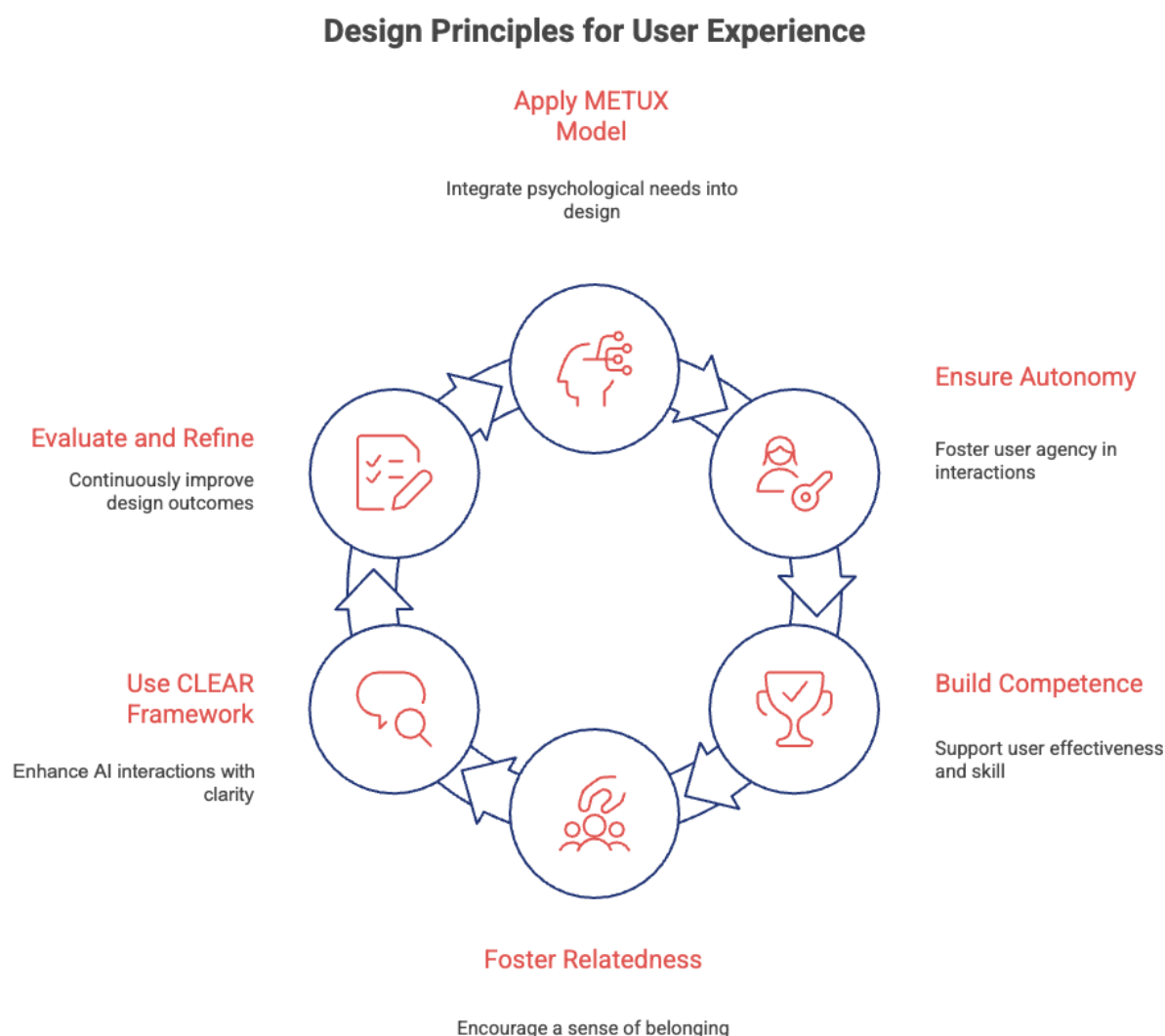


Fig. 5. Design patterns of AR and AI-powered implementations
(own figure, created with napkin.ai).

Through the variety of patterns identified, we hope to help developers, instructional designers and teachers – as well as learners, of course – to make users feel capable and effective by designing for the right level of challenge, providing clear feedback and enhancing the number of opportunities for learning and mastery. We also hope that these design features foster the development of genuine feelings of connection and belonging for the learners.

We have striven to adhere to these principles but would be glad to rectify any instances where we have fallen short in this respect at a later stage.

We predict that AI+AR will be particularly valuable in enabling more individualised learning by providing learners with readily available practice opportunities, while offering immediate feedback, enhancing opportunities for interaction and potentially increasing their willingness to communicate (see also Sahin Kizil et al., 2025²²; see Geroimenko, 2023²³).

Writing a guide this early without decades of empirical study comes at a cost. Sahin Kizil et al. (2025), for example, identify significant research gaps, particularly regarding the limited empirical evidence available as to the effectiveness of this technology as regards long-term learning outcomes. They call for future studies involving larger, more diverse sample users, the extended duration of interventions and a focus on understanding impacts across different learner characteristics, such as age, personality and teacher roles. There is also a noted gap in research exploring language learners' higher-order cognitive skills, the emotional support capabilities of AI, and effective learning design strategies for AI integration. We could not agree more.

1.4 MirageXR

MirageXR is an educational AR platform that enables experts and learners to create and experience AR content. MirageXR can use either marker-based approach to position 3D content or the marker-less approach building a so-called spatial map of the physical space around the user and uses its features as anchors for the digital content.

MirageXR supports in-situ authoring of AR content that does not require programming skills. Teachers can design AR learning activities in the MirageXR app and save them in a secure database, while the students can access these activities via the same MirageXR app.

The same application is required for both authoring and playing the AR content. MirageXR is a cross-platform tool, supporting smart phones with iOS and Android²⁴ and the Apple Vision Pro smart glasses (earlier also supported HoloLens 2 AR glasses).

The learning content created in MirageXR is organized as AR activities with multiple steps, supporting procedural training and operational support for complex tasks. The app allows creating learning content from various content types, such as animated 3D models, images, text labels, and AI-powered characters called virtual instructors. We will refer in particular to these virtual instructors further in this guide when discussing prompt engineering (Fig. 6).

²² Şahin Kızıl, A., Klimova, B., Pikhart, M., & Parmaxi, A. (2025). A systematic review of the recent research on the usefulness of chatbots for language education. *Journal of Computer Assisted Learning*, 41(2) <https://doi.org/10.1111/jcal.70001>

²³ Geroimenko, V. ed., 2023. *Augmented reality and artificial intelligence: The fusion of advanced technologies*. Springer Nature.

²⁴ iOS: <https://apps.apple.com/sr/app/miragexr/id1561154572>, Android: <https://play.google.com/store/apps/details?id=com.WEKIT.MirageXR>

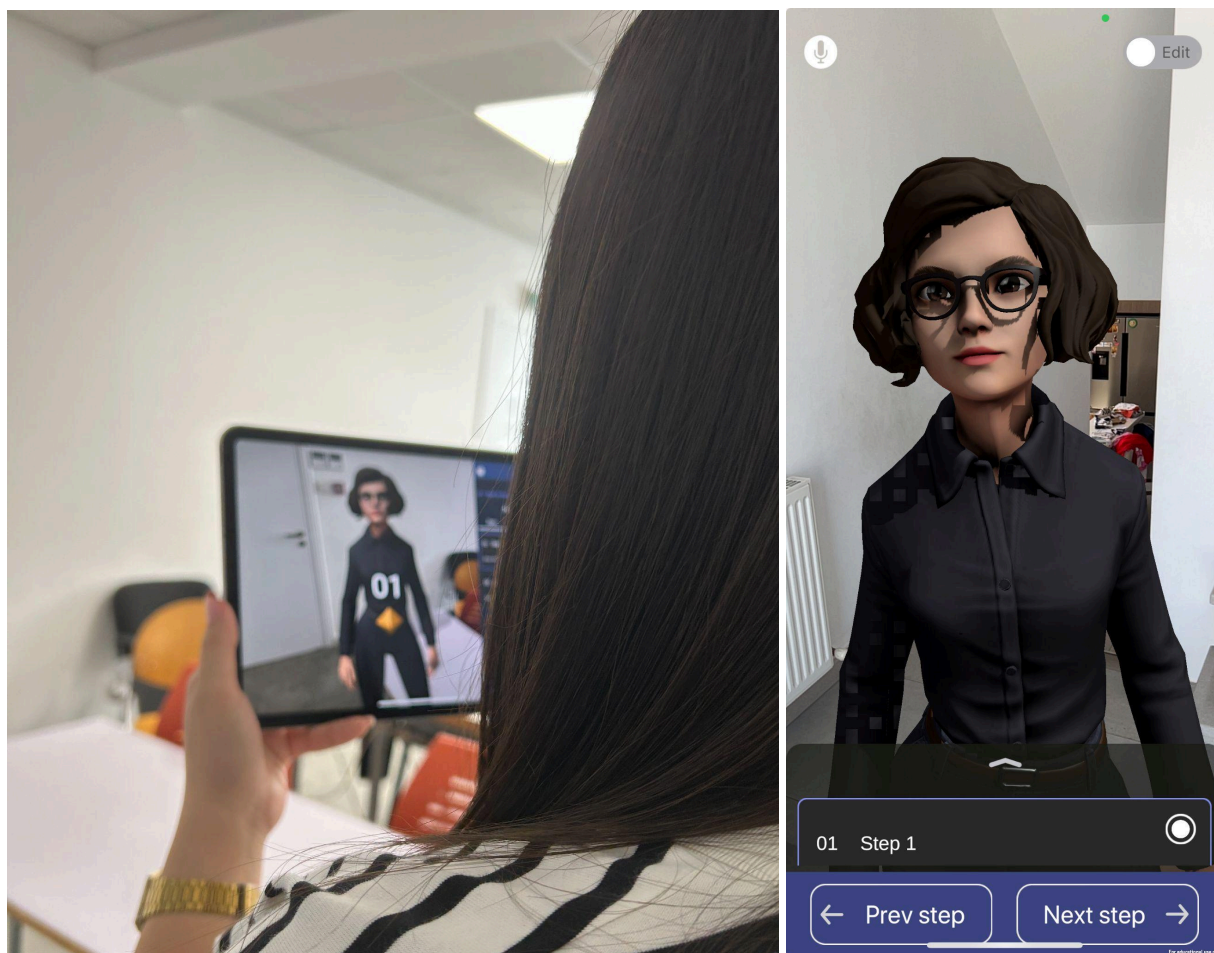


Fig. 6. A dialog with a Virtual Instructor in MirageXR

1.5 How to read this guide

This guide is organised in five sections:

- **Section 1** provides a brief introduction to prompts and their use in language learning.
- **Section 2** explains what we consider to be good practice in writing prompts and how using several of our well-matched and bounded examples as patterns²⁵ can offer others a basis for writing their own prompts.
- **Section 3** focuses on language pedagogical patterns.
- **Section 4** looks at situated task practice.
- **Section 5** summarises what we achieved in this pattern language and where we see gaps and opportunities for future work.

You can read this guide beginning to end or you can follow the index to look up patterns suitable for your specific teaching needs.

²⁵ “Prompt patterns” refer here to recurring structures and formats used when designing AI prompts in the context of instructing, e.g., ChatGPT.

2. Prompt patterns and a pattern language

To align and formalise the descriptions of prompts in a form that supports easy interchange and reuse, this guide resorts to the idea of *design patterns*.

The basic idea of a design pattern is to encapsulate the descriptions of problems and their solutions. Originally proposed by Alexander (1979)²⁶ to promote solutions in architecture, the idea of design patterns was picked up by software engineers (Gamma et al., 1994)²⁷ to promote reusable, recurring software programming solutions, before finding a foothold more broadly in education²⁸, even leading to proposals to redefine teaching as a design science²⁹. Traditionally (Alexander, 1979), design patterns are composed of: a name ('headline'), the problem description, the solution description (sometimes including a sketch) and information about the appropriate usage context. But, depending on the application area, a range of additional elements can be described, such as the following for pedagogical patterns: forces³⁰, consequences, audience, resources, example instances, contra indications³¹.

Classically, a pattern is a tested and trialled solution to a defined problem, in many cases requiring multiple independent instances to be identified to justify it. We have tried to keep to this rule as much as possible, working off larger prompt collections that we have collated, but we have not always been strict with this. In our small community, great ideas spread like wildfire and it thus seemed too high a bar to require independent discovery.

The focus of this guide is on design patterns for AI+AR prompts, modifying the format of pattern descriptions slightly to include, in a self-contained way, all information required to configure an AI+AR enabled virtual human. In a few cases, additional information might be added explaining how to set up the complementary AR scene and environment to reflect the needs of the conversational AI+AR virtual human.

As all the existing pattern methodologies we revisited (see Sobernig et al., 2006) fall short of describing the key elements required for AI+AR prompts, we propose the following key elements to be added: prompt, context and knowledgebase. These commonly serve as direct inputs to the Application Programming Interface (API) calls of Large Language Models (LLM) providers, such as OpenAI. To be clear, these three elements are not simply descriptive but they directly reflect different input parameters of the API calls sent to LLM providers during the execution of conversations with a chatbot.

²⁶ Alexander, C. (1979): The timeless way of building (Vol. 1). New York: Oxford University Press.

²⁷ Gamma, E.; Helm, R.; Johnson, R.; Vlissides, J. (1994): Design Patterns: Elements of Reusable Object-Oriented Software. Reading, Mass.: Addison-Wesley.

²⁸ Sobernig, S., Danielewska-Tulecka, A., Wild, F., Kusiak, J. (2006): Interoperability and Patterns in Technology-Enhanced Learning. Polish Information Processing Society (PTI), Szczyrk, 2006.

²⁹ Laurillard, D., 2013. Teaching as a design science: Building pedagogical patterns for learning and technology. Routledge.

³⁰ E.g., thoroughness vs limited time

³¹ Bergin, J., Eckstein, J., Volter, M., Sipos, M., Wallingford, E., Marquardt, K., Chandler, J., Sharp, H. and Manns, M.L. (2012): Pedagogical patterns: advice for educators. Joseph Bergin Software Tools.
<http://www.pedagogicalpatterns.org/>

With the help of these three inputs, we can create a persona with specific conversational traits and behaviour. We call this persona a '*Virtual Instructor*', a term that we prefer over the more general ones, such as agent, AI, chatbot, virtual human or character.

Once a Virtual Instructor is created with the help of the prompt patterns described in this guide, learners can interact with it inside MirageXR (see section 1.4), an AR+AI system available as open source and for free. The patterns, however, can be applied widely to most commercially and non-commercially available services.

2.1 The pattern format

Context descriptions, the knowledgebase and the prompt can be defined by the teacher or instructional designer when creating a learning activity that involves AI+AR Virtual Instructors.

In *MirageXR*, the context prompt can be added in the app directly by the teacher or instructional designer, or they can be provisioned as preconfigured AI models. Setting up preconfigured AI models includes writing the context descriptions, the knowledgebase, and the prompt (see the Configuration part on Fig. 7), and selecting a foundational LLM, such as GPT 3.5 or GPR 4.0. The preconfigured AI model can be seen as a complex behavior description, which can be used to create Virtual Instructors in multiple learning activities.

When working with MirageXR, these preconfigured AI models are stored in the repository and the secure AI gateway service called *lxr*. When creating a Virtual Instructor in the MirageXR app, a teacher or instructional designer can select the desired preconfigured AI model, together with other parameters, such as the language the Virtual Instructor should speak, the voice it should use, and the visual appearance (as an avatar 3D model) it should have (see the Frontend AR+AI application on Fig. 7). Once the preconfigured AI model is selected in the MirageXR app, the corresponding context, knowledgebase, and prompt will be injected into the conversations of the learner with the Virtual Instructor, using *lxr* as a gateway to an AI service provider.

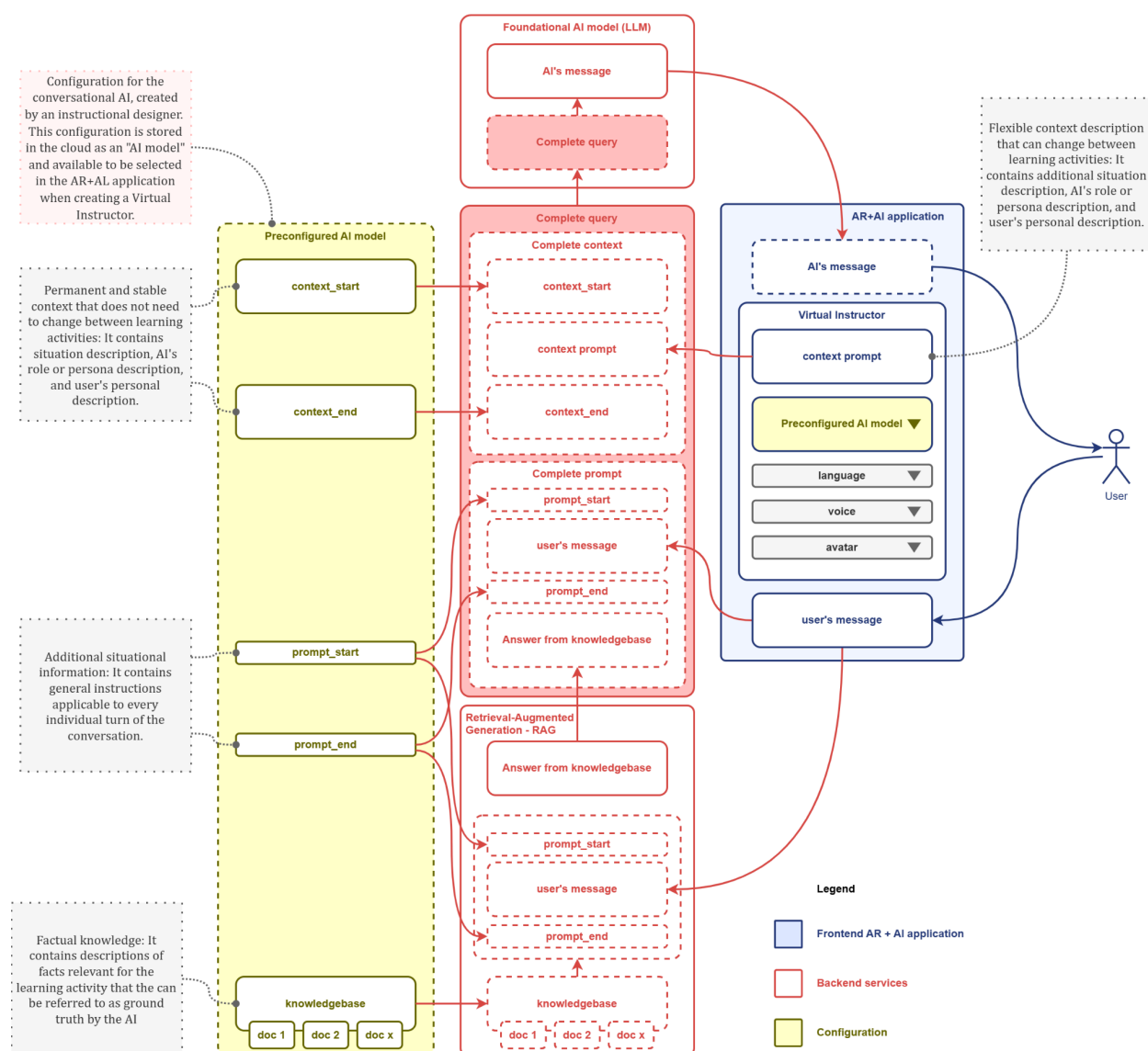


Fig. 7. Detailed workflow of interaction between the user and the AI in MirageXR AR application

2.1.1. Context

To support the design of AR-embodied AI (as virtual human instructors), the context can be understood as either the description of the situation in which the dialogue takes place or as the role the conversational partner(s) play(s) when engaging in the dialogue (their 'persona'). This context is usually defined by the teacher or the designer of the Virtual Instructor.

The first aspect of the context refers to the situation in which the dialogue takes place. It is important to describe the situation very precisely and in detail, as otherwise the dialogue can easily veer off course. Such precision also applies to the answers generated, such as the maximum word count of such responses or the need to avoid the employment of tables in a spoken exchange. For example, learners on lower language proficiency levels could benefit from word count restrictions in the AI responses, not only potentially simplifying the nature of these answers but also significantly speeding up the text-to-speech audio generation and download. We therefore

recommend adding instructions such as the following: “Restrict your answer to a maximum of 10 words”³².

The second aspect of the context refers to the personas participating in the dialogue, which can include a description of the AI persona but might also define the persona the user takes on for the duration of the conversation (e.g., in a role play).

When providing your description, you should ask yourself: ‘What describes the Virtual Instructor persona best?’ You can also describe particular mannerisms, such as how the AI responds to questions or other input from the learner (e.g., ‘Your name is Andreas and you always end your responses with ‘Right’).

Three inputs serve to configure the context:

- **context_start**: Server-sided for preconfigured models, not directly accessible by the user. Contains behaviour descriptions.
- **context prompt**: Configured in MirageXR directly in the edit panel of the Virtual Instructor. Used for situational descriptions.
- **context_end**: Server-sided for preconfigured models. Not directly accessible by the user. Contains behaviour descriptions.

This entire context provided consists of the input fields described above. It is added in each conversational turn between the user and the foundational LLM. This text string is thereby injected into the context window of the chat completion API of chatGPT using the role ‘system’ with the content ‘context’.

Note that the size of the context input field varies from AI provider to AI provider. With OpenAI’s chatGPT³³ models, the size is about 3,000 to 6,000 words. Other LLMs vary, some allowing for very large context windows.

2.1.2. Knowledgebase

Background knowledge for Retrieval-Augmented Generation or RAG (see 3.4) can be provided separately, using a knowledgebase. A knowledgebase consists of a collection of text files, which contain the factual knowledge to which the AI model of the Virtual Instructor can refer to when providing the answers to the user.

Useful sources include partial or entire textbooks or lecture notes, scientific articles, websites, databases, relevant messages, emails, etc. They are converted from their original format to plain text and then automatically split into smaller chunks. The collection as a whole needs to be relevant to the learning activity and cover the area of expertise attributed to the Virtual Instructor.

2.1.3. The user prompt (also known as the message)

When talking to the AI+AR Virtual Instructor, user speech input is transcribed and sent together with additional prompts to the AI. These additional, ‘situational’ context prompts are added to each conversational turn, when the AI response is generated, and are defined in the model entry in the AI gateway, entered by the administrator during model provision.

³² Or more: A1 = 10 words, A2 = 15 words, B1 = 20, B2 = 25, C1 = 30, C2 = 35, as a rule of thumb

³³ <https://platform.openai.com/docs/models>

For example, this might include instructions to the virtual instructor such as ‘Please respond in German!’.

The structure of the prompt is the following³⁴:

- **prompt_start**: Inserted before the user dialogue message.
- **prompt**: the current dialogue move from the user (aka the user’s ‘message’)
- **prompt_end**: Inserted after the user dialogue message.

2.2 A prompt engineering template

We recommend that you use the following template, which addresses all aspects relevant to the design of Virtual Instructor prompts.

While the sections (or ‘fields’) do not match directly on those mentioned above (i.e., system context, knowledgebase, prompt), they will be conflated to them when the system communication takes place at the end, APIs for AI communication only allow for a very limited number of fields.

1. Name or identifier of the prompt: Give your prompt a clear name or identifying title.

Example: Job Interview at Google

2. Context description: Describe the virtual instructor persona and the setting of the conversation. This enables the virtual instructor to understand the situation.

(*@context_prompt* or *@context_start*)

Example: You are Maria, a friendly and skilled HR manager at Google’s Cyprus office. You need to interview a candidate for an engineering role. You are friendly and professional, ask open-ended questions about motivation, interest, and challenges.

3. Instructions for interaction: Explain how the virtual instructor should behave during the conversation. Include how it should start, respond, and end the interaction. (*@context_end*)

Example: Start the interview with a relaxed introduction to the job and responsibilities. Ask one question at a time and wait for the user’s answer. End the conversation after five questions and give feedback on their use of language.

4. User profile: Briefly describe the user who will be interacting with the virtual instructor (e.g., language level, background). (*@context_start* after 2)

Example: The user is a language learner of English at B2 level.

5. Knowledgebase: List any specific knowledge or data that the AI should draw on (by including into a knowledgebase) or ignore (by not including in a knowledgebase). (either *@context_end* or *@knowledgebase*)

Example:

³⁴ More technically described: with every utterance, the AI receives a concatenated string of *prompt_start* + text received from speech-to-text service of the user’s utterance + *prompt_end*.

The job is in engineering at Google's Cyprus office.
Key responsibilities: teamwork, software design, problem-solving, and clear communication.
The ideal candidate is collaborative, motivated, and eager to learn.

6. Language and style guidelines: Specify the language level, tone, length, and style of the AI's responses. Include any special instructions (e.g., no emojis, no bullet points). (*@prompt_start* or *@prompt_end* to avoid override by user input)

Example: Keep answers under 30 words in length. Correct mistakes gently and politely. Avoid emojis and bullet points. Adapt language to a B2-level English learner. Use simple language.

2.3 Pattern language

Design patterns that complement one other harmonically form a design pattern language. This guide is our attempt to provide such pattern language for the design of AI+AR Virtual Instructors. While still incomplete, it offers a coherent collection of useful templates. The subsequent Sections 4 and 5 provide the design patterns of this pattern language.

3. Technical considerations

Prompt engineering is the art and science of designing effective inputs for LLMs to guide their outputs toward desired behaviors. Technically, this involves understanding not only the capabilities and limitations of LLMs, but also how variations in prompt design influence model reasoning, performance and reliability. Prompt engineering is foundational, affecting everything from task accuracy to interpretability. But it is also an evolving field and we anticipate further advances leveraged by technical innovation and cognitive insights. Competent prompt design is key to the safe and efficient use of AI.

Sahoo et al. (2025)³⁵ provide a comprehensive overview, identifying 41 distinct prompt engineering techniques, including zero-shot, few-shot, and chain-of-thought prompting as well as advanced strategies like self-consistency, RAG, and prompt fine-tuning.

The following subsections provide a more detailed description of what we consider to be core techniques, along with pointers to key papers that may be useful for readers wishing to gain a deeper understanding of them.

3.1 Chain of Thought

Chain of Thought (CoT) prompting is an innovative technique for enhancing how AI models solve complex problems. It is particularly useful in educational contexts (Wei et al., 2022³⁶; Wei & Zhou, 2022³⁷). Traditional AI prompting asks a model directly for an answer to a question. In contrast, CoT prompting instructs AI models to explicitly outline intermediate reasoning steps, mirroring how humans naturally break down complicated problems into simpler stages.

For example, rather than merely providing the solution to a maths word problem, the AI would first explicitly write down its intermediate reasoning, calculating and articulating each step before arriving at the final answer.

This approach significantly improves the performance of AI, enabling language models to perform better by breaking down complex problems into more manageable parts. The intermediate steps are shown to the user, thus clarifying how the answers were derived.

Self-consistency, according to Wang et al. (2023)³⁸, can significantly further enhance CoT reasoning by overcoming the limitations associated with the typical greedy decoding approach. Greedy decoding is a method used in language models where, at each step of generating text, the word with the highest predicted probability is chosen immediately, without considering alternative possibilities and future predictions. With a self-consistency strategy, instead of selecting a single

³⁵ Sahoo, P., Singh, A.K., Saha, S., Jain, V., Mondal, S. and Chadha, A. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications. <https://doi.org/10.48550/arXiv.2402.07927>.

³⁶ Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V. and Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35. <https://dl.acm.org/doi/10.5555/3600270.3602070>.

³⁷ Wei, J.; Zhou, D. (2022). Language Models Perform Reasoning via Chain of Thought, Online at: <https://research.google/blog/language-models-perform-reasoning-via-chain-of-thought/>

³⁸ Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A. and Zhou, D., 2022. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.

reasoning path (as in greedy decoding), a diverse set of potential reasoning paths is sampled. Each sampled reasoning path independently arrives at an answer. After generating these multiple reasoning paths, self-consistency aggregates the final answers by selecting the most frequently occurring one. This method leverages the principle that correct reasoning processes, although diverse, usually converge on the same correct solution. This reduces the susceptibility to errors arising from individual reasoning steps.

When designing prompts for an AI+AR language learning application, it is important to structure prompts that encourage the step-by-step reasoning of CoT. This improves the interpretability and reliability of responses by guiding the model to break down complex tasks into smaller, logical steps before producing the answer. This approach not only enhances the quality of the AI output but also aligns with pedagogical goals by modelling the cognitive processes that learners are expected to develop themselves. For example, instead of asking an AI assistant in an AR language learning app to: “Correct any unidiomatic expressions”, a CoT-style prompt might be: “Let’s go through this step by step. First, identify any unidiomatic expressions. Then, act as an academic editor and rephrase the sentence to enhance clarity and maintain a professional tone”.

Research has shown that CoT prompting improves accuracy in reasoning tasks (Wei et al., 2022), and aligns well with constructivist learning principles by making thinking visible (Chi & Wylie, 2014)³⁹.

3.2 Few Shot (FS), One-shot (1S), Zero-shot (0S)

Retraining and fine-tuning LLMs is costly, lengthy and energy consuming. Retraining means changing the document corpus from which the LLM is created. Fine-tuning involves updating an existing, pre-trained model's internal parameters by training it on a large, task-specific dataset (e.g., thousands of examples of language exercises and learner answers).

Brown et al. (2020)⁴⁰ were able to show that, by providing examples, models can be adapted almost instantly, avoiding the huge computational cost of retraining and fine-tuning.

Instead of training and fine-tuning the model, Few Shot (FS) prompts use a small number of examples (typically 2–10), provided directly in the prompt. The AI then learns the task instantly without modifying internal parameters. A specific case of FS prompting is where the model receives just one example in the prompt to learn from immediately, called ‘One-Shot (1S)’. Sometimes cases where no example is provided and the AI works directly only with a prompt are called Zero-Shot (0S)’.

3.3 Multitask learning

Explicit multitask learning contributes to better LLMs by directly training them on a variety of clearly defined tasks, rather than relying solely on general text prediction, according to Sanh et al.

³⁹ Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, 49(4). <https://doi.org/10.1080/00461520.2014.965823>.

⁴⁰ Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. and Agarwal, S. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33. <https://dl.acm.org/doi/abs/10.5555/3495724.3495883>.

(2022)⁴¹. This method improves an LLM's ability to generalise, enabling it to perform well even on tasks it has not explicitly been trained on, a capability known as zero-shot generalization. The training utilises a range of diverse tasks such as question-answering, summarization, and sentiment analysis. Each task is represented using clearly defined prompts, guiding the model explicitly about what the task is and how to approach it. Training involves datasets converted into prompts with varied wording, enhancing the model's adaptability.

This helps achieve improved Zero-Shot generalization, where the trained LLMs become more capable of accurately handling tasks they have never directly encountered, because they recognize task structures and instructions from similar previous training. The training on multiple prompts per task makes LLMs less sensitive to variations in how instructions are worded, improving their robustness to differing prompt input. In consequence, models achieve better performance on diverse tasks without necessarily becoming significantly larger, outperforming even much larger models trained only on general language data.

3.4 Retrieval-Augmented Generation

Retrieval-Augmented Generation (RAG) enhances language models by combining parametric memory (pre-trained sequence-to-sequence language models) with non-parametric memory (retrieval-based knowledge, typically from a knowledgebase like Wikipedia) (see Lewis et al. , 2021)⁴². It operates by first encoding an input query, which retrieves relevant passages from an indexed knowledge database and then generating answers conditioned on these retrieved documents. RAG models use latent document marginalisation, considering multiple retrieved passages to produce accurate outputs, either by relying on one document per generated sequence or allowing each token to be conditioned on different documents.

In simple terms, RAG works by searching an existing, AI-external document space (e.g., documents provided by a prompt designer in the knowledgebase) for relevant information regarding a question asked by the user. Instead of relying only on the LLM's knowledge, RAG automatically looks through the knowledgebase to identify those documents (or document parts) that closely match the user's prompt. Typically, the comparison is done using similarity measures between the vector embeddings rather than the original texts, but in principle any search technology can be applied here to retrieve the document or website containing the most likely answer. Once the relevant document or document part is identified with this search, its content is appended to the context field of the call to the LLM to create a well-formulated answer.

By retrieving relevant information during generation, RAG produces more accurate, factual and diverse responses. It can help avoid hallucinations, leveraging externally verified sources to decrease incorrect or fabricated outputs. External memory can also more easily and more flexibly be updated without retraining the entire model. In addition, the retrieved documents provide more transparency regarding the sources used by the model to create answers.

⁴¹ Sanh, V., Webson, A., Raffel, C., Bach, S.H., Sutawika, L., Alyafeai, Z., Chaffin, A., Stiegler, A., Scao, T.L., Raja, A. and Dey, M. (2021) Multitask prompted training enables zero-shot task generalization. <https://doi.org/10.48550/arXiv.2110.08207>.

⁴² Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T. and Riedel, S. (2020) Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33. <https://dl.acm.org/doi/abs/10.5555/3495724.3496517>.

3.5 Memory

Underlying LLM architectures (i.e., transformers) are effectively stateless: from one call to the next, they do not remember, which is their big advantage in respect to processing. But this means to establish a real conversation where each turn builds on the previous that such conversational memory must be passed in each call using the context prompt.

Prompts typically include a summary of the prior dialogue or the prior dialogue in full to help maintain continuity in the conversation. This is especially important in longer conversations that unfold over time or across phases, where the learner and AI need a shared memory of what has already occurred.

This is usually handled by the underlying system or API. Nevertheless, explicit reference to and control over the conversation history or the prompt commands referring to it can provide a powerful instrument for steering the dialogue. This can leverage cognitive processes such as scaffolding and fading (Gonulal & Loewen, 2018⁴³).

This technique aligns with research on contextualized learning and discourse coherence, which emphasizes the role of prior knowledge and memory cues in supporting comprehension and learning (see Chi, 2009⁴⁴).

3.6 Fine-tuning

A certain degree of fine-tuning can still be conducted when using the LLM, not requiring the training of a new model with new inputs. Familiarisation with these advanced aspects can help to further optimise app experiences (and associated costs). The following parameters can be controlled by developers at runtime to influence how the model responds⁴⁵:

- The **temperature** value adjusts how random or creative the output is. Higher values make the responses more varied, while lower values make them more focused and deterministic. For example, a temperature value closer to 0 will result in more predictable text. This could be useful for language instructors who want the AI to generate responses that are very commonly used.
A temperature value closer to 1 will result in more random text. This could be useful for language instructors who want the AI to generate more creative and open-ended responses. For example, if a student was focusing on their creative writing, the instructor could set the temperature value to a higher level to encourage the AI to generate more unique conversations.
- The **top_p** parameter is another way to control randomness, by limiting the choices to the most probable tokens that together reach a certain probability mass. Temperature or top_p can be used alternatively, but not both.

⁴³ Gonulal, T. and Loewen, S., 2018. Scaffolding technique. The TESOL encyclopedia of English language teaching, pp.1-5.

⁴⁴ Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. Topics in cognitive science, 1(1). <https://doi.org/10.1111/j.1756-8765.2008.01005.x>.

⁴⁵ <https://platform.openai.com/docs/api-reference/introduction>

- The *max_tokens* parameter limits the length of the model's response. The *presence_penalty* discourages the model from repeating concepts that have already appeared, encouraging it to explore new ideas instead, while the *frequency_penalty* discourages the model from repeating the same words or phrases too often.
- *Stop* sequences can be used to tell the model when to stop generating further text.

3.7 Restrictions arising from web communication

The technical implementation of an AI-enabled system can vary, and the differences between streaming and request-based systems, for example, can be significant in terms of their user experience. *Request-based* means that, with every turn, the communicating application builds up a connection to the server, sending first the audio recording to receive a transcript, which is then injected as a message into a chat completion thread. *Streaming* communication differs from this as it seeks to transmit audio directly while recording, starting transcription already with the first word to speed up the process, while adding complexity to the software architecture needed.

By design, request-based systems suffer from slow response times, primarily caused by the limitations of synchronous processing, which requires full recording and upload / full generation and download of any audio files before their playback. At the same time, request-based systems are easier to implement than streaming services, hence their prevalence in the market.

For a Virtual Instructor, several AI services are chained together to create the intended results. First, audio input is recorded from the user of the system. This is then transcribed to text using speech-to-text services like OpenAI Whisper, IBM Watson or Google Speech-To-Text. Next, this text is sent to the LLM chat completion API to trigger a response in the conversation. Finally, a text-to-speech service is used to generate the audio output to be played to the user. Typically, in an AR system, this is accompanied by additional services for generating avatar body animations, gestures, facial expressions and lip movement, with varying degrees of complexity. Sometimes the Virtual Instructor additionally sends data with control information about this or even sends events to trigger the execution of specific functionality within the app.

Direct, synchronous web requests are easy to implement, while web socket connections / webRTC are required for streaming communication. The big difference between these is merely in speech in and output, as synchronous requests require user speech recording / speech output generation to be concluded before upload, and upload to be concluded before processing. Streaming starts with processing already during transmission of any audio content.

4. Language related patterns

Each subsection in this section (and the following two) describes an individual pattern. Each pattern is presented with a description of the problem it solves, the usage context, and the proposed solution, with examples highlighted in “dark grey”.

Prompting is a little bit like cooking. Only the combination of ingredients makes great dishes, but sometimes less is more. Different to cookery, though, prompt engineering benefits from (and is conducive to) computational thinking: it is most powerful when we formulate instructions in natural language in a way that makes it easier for a computer to understand and execute.

Each prompt part has a clear role, just like in a computer program, and only the set of instructions together achieves the desired effect.

Consider this:

| Prompt to the AI | Instruction |
|--|---------------|
| “Teach me about the second world war.” | Set topic |
| “Do a quiz with me and ask me five surprising questions that can be answered only with yes or no.” | Set task |
| “Ask questions one after the other, and wait for my answers, before you ask the next question.” | Set style |
| “Once you have collected all my answers, provide feedback to me.” | End condition |

The column on the right highlights what each (natural language) instruction focuses on. Try it, the result is convincing, but also demonstrates that additional prompts will be needed, when using such prompts to animate an AR+AI Virtual Instructor. For example, the feedback instruction creates enumerations and summary scores that are difficult to read out with text-to-speech (e.g., “3/5” might be pronounced “three divided by five” or “three over five”, not “three out of five”).

To assist with the nuances in design, we have extracted design patterns from a prompt collection.

In this section, we have grouped together the design patterns for prompts that have primarily to do with the use of language – choosing the prompt language (4.1), setting proficiency levels (4.2), or addressing specific elements of language (4.3). Moreover, we illustrate how LLM can support multimodal learning through multimodal input, i.e. text, images, sound (4.4).

4.1 Prompt language

The language of the prompt itself affects system behavior. This is especially true for prompts that are written in any language other than English. Mondshine et al. (2025)⁴⁶ found in their work that LLMs provide better results in the language they were trained in, which is usually English.

⁴⁶ Mondshine, I., Paz-Argaman, T. and Tsarfaty, R. (2025) Beyond English: The impact of prompt translation strategies across languages and tasks in multilingual LLMs. <https://doi.org/10.18653/v1/2025.loresmt-1.9>.

Improved results can be expected when the prompt is written in a high-resource language. This refers to languages for which a large amount of digital material and resources are available, which is usually English. Such resources can be, for instance, large annotated corpora, speech and audio data sets, text-to-speech (TTS) and automatic speech recognition (ASR) systems.

4.2 Language proficiency levels (from A1 to C2)

A competent speaker of a language will have to move in instructed language acquisition through proficiency levels as described, e.g., in the Common European Framework of References (CEFR). Much has been written about these on the Internet, and consequently, they are no secret to LLMs, and prompts can refer to these language levels (Fig. 7).

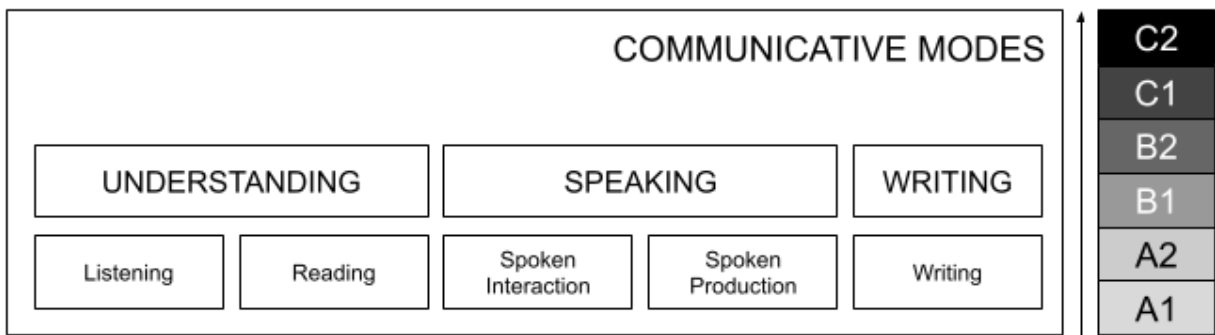


Fig. 7. CEFR levels.

For example, the context prompt of an AI+AR Virtual Instructor could benefit this way from the following input: “Use popular German phrases in your answers on language learning level B2. Congratulate me on my language use whenever I use a popular German phrase.”

Figure 8 provides insights on what such phraseology prompts can generate, illustrating phrase lists produced by chatGPT for specific language levels.

| | |
|---|---|
| <p>list the 20 most popular german phrases</p> <p>Hier sind 20 der beliebtesten deutschen Redewendungen – ideal für B2-Lernende:</p> <p>Allgemeine Ausdrücke:</p> <ol style="list-style-type: none"> 1. Alles klar! – <i>(All right! / Got it!)</i> 2. Lass uns loslegen! – <i>(Let's get started!)</i> 3. Es kommt darauf an. – <i>(It depends.)</i> 4. So ist das Leben. – <i>(That's life.)</i> 5. Ich habe die Nase voll! – <i>(I've had enough!)</i> <p>Zustimmung und Meinungen:</p> <ol style="list-style-type: none"> 6. Ich bin ganz deiner Meinung. – <i>(I completely agree with you.)</i> 7. Lassen Sie uns zum Punkt kommen. – <i>(Let's get to the point.)</i> 8. Das ist leichter gesagt als getan. – <i>(That's easier said than done.)</i> 9. Dafür gibt es keine Entschuldigung. – <i>(There's no excuse for that.)</i> 10. Das macht Sinn. – <i>(That makes sense.)</i> <p>Alltag & Emotionen:</p> <ol style="list-style-type: none"> 11. Ich stehe auf dem Schlauch. – <i>(I'm confused.)</i> 12. Hast du etwas auf dem Herzen? – <i>(Do you have something on your mind?)</i> 13. Ich drücke dir die Daumen! – <i>(I'll keep my fingers crossed for you!)</i> 14. Das ist nicht mein Bier. – <i>(That's none of my business.)</i> 15. Ich verstehe nur Bahnhof. – <i>(I don't understand anything.)</i> <p>Reisen & Höflichkeit:</p> <ol style="list-style-type: none"> 16. Es war mir eine Freude. – <i>(It was my pleasure.)</i> 17. Gute Reise! – <i>(Have a good trip!)</i> 18. Kann ich mit Karte zahlen? – <i>(Can I pay by card?)</i> 19. Wie komme ich am besten nach ...? – <i>(What's the best way to get to ...?)</i> 20. Könnten Sie das bitte wiederholen? – <i>(Could you please repeat that?)</i> | <p>list the 20 most popular german phrases for language level A1</p> <p>Hier sind 20 der beliebtesten deutschen Phrasen für A1-Lernende, die dir im Alltag helfen werden:</p> <p>Grundlegende Begrüßungen & Verabschiedungen:</p> <ol style="list-style-type: none"> 1. Hallo! – <i>(Hello!)</i> 2. Guten Morgen! – <i>(Good morning!)</i> 3. Guten Abend! – <i>(Good evening!)</i> 4. Tschüss! – <i>(Bye!)</i> 5. Auf Wiedersehen! – <i>(Goodbye!)</i> <p>Höfliche Ausdrücke:</p> <ol style="list-style-type: none"> 6. Bitte. – <i>(Please.)</i> 7. Danke! – <i>(Thank you!)</i> 8. Vielen Dank! – <i>(Thank you very much!)</i> 9. Gern geschehen! – <i>(You're welcome!)</i> 10. Entschuldigung! – <i>(Excuse me! / Sorry!)</i> <p>Vorstellen & Fragen nach Namen:</p> <ol style="list-style-type: none"> 11. Wie heißt du? – <i>(What's your name?)</i> 12. Ich heiße ... – <i>(My name is ...)</i> 13. Freut mich! – <i>(Nice to meet you!)</i> 14. Woher kommst du? – <i>(Where are you from?)</i> 15. Ich komme aus ... – <i>(I come from ...)</i> <p>Alltägliche Fragen & Antworten:</p> <ol style="list-style-type: none"> 16. Wie geht es dir? – <i>(How are you?)</i> 17. Mir geht es gut, danke. – <i>(I'm fine, thank you.)</i> 18. Können Sie das bitte wiederholen? – <i>(Can you please repeat that?)</i> 19. Ich verstehe nicht. – <i>(I don't understand.)</i> 20. Sprechen Sie Englisch? – <i>(Do you speak English?)</i> |
| <p>list the 20 most popular german phrases for language level B2</p> <p>Hier sind 20 beliebte deutsche Phrasen für das B2-Niveau, die dir helfen, fortgeschrittene Gespräche zu führen:</p> <p>Meinung ausdrücken & Diskussionen:</p> <ol style="list-style-type: none"> 1. Meiner Meinung nach ... – <i>(In my opinion ...)</i> 2. Ich bin ganz deiner Meinung. – <i>(I completely agree with you.)</i> 3. Das sehe ich anders. – <i>(I see it differently.)</i> 4. Es kommt darauf an. – <i>(It depends.)</i> 5. Lassen Sie uns zum Punkt kommen. – <i>(Let's get to the point.)</i> <p>Alltag & Redewendungen:</p> <ol style="list-style-type: none"> 6. Das ist leichter gesagt als getan. – <i>(That's easier said than done.)</i> 7. Ich stehe auf dem Schlauch. – <i>(I'm confused.)</i> 8. Ich habe das im Griff. – <i>(I have it under control.)</i> 9. So kann es nicht weitergehen. – <i>(It can't go on like this.)</i> 10. Dafür gibt es keine Entschuldigung. – <i>(There's no excuse for that.)</i> <p>Gefühle & Reaktionen:</p> <ol style="list-style-type: none"> 11. Das ist mir völlig egal. – <i>(I don't care at all.)</i> 12. Das war mir nicht bewusst. – <i>(I wasn't aware of that.)</i> 13. Hast du etwas auf dem Herzen? – <i>(Do you have something on your mind?)</i> 14. Das hat mich umgehauen! – <i>(That blew me away!)</i> 15. Ich bin mit den Nerven am Ende. – <i>(I'm at my wit's end.)</i> <p>Probleme & Lösungen:</p> <ol style="list-style-type: none"> 16. Wie soll ich das verstehen? – <i>(How should I understand that?)</i> 17. Das macht keinen Sinn. – <i>(That doesn't make sense.)</i> 18. Es war mir eine Freude. – <i>(It was my pleasure.)</i> 19. Das ist nicht mein Bier. – <i>(That's none of my business.)</i> | <p>list the 20 most popular german phrases for language level C2</p> <p>Hier sind 20 beliebte deutsche Phrasen für das C2-Niveau, die dir helfen, auf muttersprachlichem Niveau zu kommunizieren:</p> <p>Fortgeschrittene Meinungsäußerung & Argumentation:</p> <ol style="list-style-type: none"> 1. Es liegt auf der Hand, dass ... – <i>(It is obvious that ...)</i> 2. Man kann nicht leugnen, dass ... – <i>(One cannot deny that ...)</i> 3. Unter der Voraussetzung, dass ... – <i>(Provided that ...)</i> 4. Es lässt sich nicht bestreiten, dass ... – <i>(It cannot be disputed that ...)</i> 5. Das ist ein zweischneidiges Schwert. – <i>(That's a double-edged sword.)</i> <p>Redewendungen & gehobene Sprache:</p> <ol style="list-style-type: none"> 6. Da beißt die Maus keinen Faden ab. – <i>(That's a fact; it can't be changed.)</i> 7. Das ist Schnee von gestern. – <i>(That's old news.)</i> 8. Den Nagel auf den Kopf treffen. – <i>(To hit the nail on the head.)</i> 9. Ich lasse mich nicht aufs Glatteis führen. – <i>(I won't be tricked.)</i> 10. Das ist ein Tropfen auf den heißen Stein. – <i>(That's a drop in the ocean.)</i> <p>Nuancierte Emotionen & Reaktionen:</p> <ol style="list-style-type: none"> 11. Mir fehlen die Worte. – <i>(I am speechless.)</i> 12. Das lässt tief blicken. – <i>(That speaks volumes.)</i> 13. Das ist nicht von der Hand zu weisen. – <i>(That cannot be dismissed.)</i> 14. Ich nehme das mit einem Augenzwinkern. – <i>(I take that with a grain of salt.)</i> 15. Das geht mir gegen den Strich. – <i>(That rubs me the wrong way.)</i> <p>Komplexe Problemlösungen & Reflexion:</p> <ol style="list-style-type: none"> 16. Das bedarf einer genaueren Betrachtung. – <i>(That requires a closer look.)</i> 17. Daraus kann man schließen, dass ... – <i>(From this, one can conclude that ...)</i> 18. Es wäre vermessen zu behaupten, dass ... – <i>(It would be presumptuous to claim that ...)</i> 19. Das führt unweigerlich zu der Frage ... – <i>(That inevitably leads to the question ...)</i> 20. Das ist eine Gratwanderung. – <i>(That's a fine line.)</i> |

Fig. 8. Example phrase lists created by relevant prompts.

4.3 Focusing on specific language elements

Elements of language (see Parmaxi et. al, 2025⁴⁷) refer to: vocabulary, pronunciation, spelling, morphology, grammar, phraseology, meaning, and rhetoric and culture. In Section 4.3, we explain this in more detail.

Language has many aspects, which are commonly divided into different linguistic branches in their study. Our understanding of common language elements are listed in Table 1, with illustrating example prompts added for each.

| Element | Description | Example prompt |
|--------------------|---|--|
| <i>Phonetics</i> | Study of the speech facility, acoustics | Handled on text-to-speech output level: instruction prompt for, e.g., Whisper text-to-speech: “Increase speed gradually throughout the text.” |
| <i>Phonology</i> | Study of how speech sounds or the corresponding units in signed languages are categorised and organized in a language | “You speak German with a Saxonian dialect.” “Use a British accent throughout.” |
| <i>Orthography</i> | Spelling-level; study of the spelling system of a language | “Whenever I misspell something, please correct me.” |
| <i>Morphology</i> | Word-level; study of morphemes, lexemes and compounding | “Whenever I use the wrong word ending, please correct me.” |
| <i>Syntax</i> | Sentence structure and formation, rules-level, grammar | “Please point out any grammatical mistakes I make and explain to me how to correct them.” |
| <i>Semantics</i> | Study of meaning | “I am learning German on level A1. You speak only German with me. You use only the 100 most frequent words in your answers.” |
| <i>Pragmatics</i> | Culture- and context-level: study of discourse, rhetorical structure | “Use popular German phrases in your answers.” |

Table 1. Common linguistic elements relevant for language learning.

To become a competent speaker of a language, the learner must focus on all linguistic elements and develop knowledge and skills in each of them.

⁴⁷ Antigoni Parmaxi, Anke Berns, Lina Adinolfi, Alice Gruber, Mikhail Fominykh, Angeliki Voreopoulou, Fridolin Wild, Paraskevi Vassiliou, Eirini Christou, Concepción Valero-Franco, Tormod Aagaard, Stella Hadjistassou, and Pedro Jesús Molina Muñoz (2025) Instructional design guide for Augmented Reality in language learning. ARIDLL consortium. <https://aridll.eu/>

4.3.1 Phonetics

Phonetics are technically dealt with in a different API call, using text-to-speech and speech-to-text services. Both can have parameters. For example, speech generated can be set to be of fast or slow speed. Most often, from a phonetics point of view, it is the voice that differs - female voices sound different than male voices, and these have to be configured accordingly for AI+AR instructors. In MirageXR, a set of standard voices is provided, but these can be extended or even personalised, using voice cloning.

Nevertheless, multimodal capabilities of conversational AIs can support phonetics, also analysing resulting phonology, for instance, to provide feedback on aspects such as pronunciation.

Analysis is much more complicated, as phonetic features lead to identification of meaning-carrying units, i.e., are phonological. Say, for example, the system recognises a certain acoustic pattern [i], but the context suggests it should be a different acoustic pattern [y]. The system can give different kinds of feedback depending on prompt and context. For example, this could result in: “Oh, Could you repeat that? I think you meant [sy:k], not [si:k], is that right?”. Or it could provide articulatory explicit feedback: “What you want to say is [sy:k]. You have to round and protrude your lips as if you are kissing. Try.”.

4.3.2 Phonology

Phonology looks at phonemes, phonotactics, tone, and prosody. Phonemes are the smallest sound units that distinguish meaning (e.g., /p/ vs /b/ in pat vs bat). Phonotactics refer to the rules on how phonemes can be combined in a language.

For example, the sound /ŋ/ is not permitted in word-initial position in English. Spanish does not permit the consonant cluster /sp/ word-initially. In Mandarin very few consonants are acceptable in coda position/syllable endings, and consonant clusters like [spr] or [sk] are not acceptable in any position.

Tone refers to the pitch variation that can change word meaning in some languages. Mandarin Chinese, for example, is a tone language, where different heights of tone carry different meanings.

Prosody refers to the rhythm, stress, and intonation patterns in speech that affect meaning and emotion.

Language learners commonly transfer some of the phonological features from their dominant language into the new one, sounding similar to a dialect, just a Spanish-Germans dialect, for example.

Language learners often struggle with understanding dialects that differ from the standardized official variety/language often taught in schools and language courses. Dialect is of course wider than just phonology, also including phonetic, morphological, and lexical characteristics/traits.

The prompt “You speak German with a saxonian dialect.” for both text-to-speech instruction and the conversational message to the AI can help simulate dialects, as evident in the example below:

you speak german with a saxonian dialect.

Nu, da hams mir's! 😊 **Willkommen in Sachsen!** Hier redn mir a weng anders, aber das macht's doch erst scheen, od'r? **Lass uns mal loslegn!**

Wenn de was wissen wills, nur raus damit! **Es kommt druff an**, ob's um Sprache, Kultur oder e gude Empfehlung für'n Eierschecke geht. 😊 **Ich bin ganz deiner Meinung**, dass's mit'n bissl Humor gleich viel besser läuft!

Na, hoscht Lust auf 'ne kleene Unterhaltung im Sächs'schen? **Mir machn das scho!** 😊

📄 🍌 🗣️ 🔊 🎧 🔍 ↻

Text-to-speech services like OpenAI Whisper accept an 'instructions' parameter, which can be used to add prompts, such as the ones listed below in Table 2:

| Tone and emotion | Speed and pauses | Pronunciation and accent |
|--|---|--|
| "Speak softly, with a reassuring and calm voice." | "Speak slowly and clearly, pausing slightly between sentences." | "Pronounce technical terms slowly and carefully" |
| "Make the voice sound enthusiastic and energetic." | "Increase speed gradually throughout the text." | "Use a British accent throughout." |
| "Use a neutral, professional tone." | "Pause distinctly after each comma." | "Emphasize the pronunciation of foreign names." |
| Voice styling and persona | Pitch and emphasis | Volume and clarity |
| "Speak as if you are a narrator of a children's story." | "Lower your pitch at the end of questions." | "Maintain consistent loudness throughout." |
| "Use a warm, friendly, conversational style." | "Increase emphasis on keywords and technical phrases." | "Speak clearly, articulating each word distinctly." |
| "Speak formally, as if delivering an official announcement." | "Speak with varied pitch to maintain listener interest." | "Lower your volume when indicating confidentiality." |

Table 2. Voice styling with instructions to text-to-speech generation.

Speech generation requires tweaking bespoke trained voices, which are not necessarily part of the standard speech generation toolkits. AI voice cloning services such as those from Elevenlabs⁴⁸ can successfully create very authentic dialects, and provide a good selection of authentic dialects out of the box.

⁴⁸ <https://elevenlabs.io/>

Open-source text-to-speech (TTS) models have become increasingly powerful and accessible in recent years, see Xu (2025)⁴⁹. Unlike proprietary solutions, open-source TTS give full flexibility, better data privacy, and community-driven innovation. Best-in-class XTTS-v2, according to Xu, can clone a speaker's voice in multiple languages using only a 6-second sample, supports 17 languages, and captures emotion and style transfer. Although the original company ceased operations in early 2024, XTTS-v2 continues to evolve as a highly downloaded open-source project. Other prominent systems include Tacotron-based tools, FastSpeech-2, and diffusion-based synthesizers. They all have different trade-offs regarding quality, speed, multilinguality, and fine-tuning complexity.

Yamauchi et al. (2024)⁵⁰ present a novel TTS model for cross-dialect synthesis specifically aimed at pitch-accent languages such as Japanese. Pitch-accent languages use distinct pitch patterns (accents) to differentiate meaning. This model synthesizes speech in dialects different from the speaker's native accent, an essential capability for creating voice agents that communicate naturally across regions.

Unlike traditional TTS systems, which depend heavily on manually created accent dictionaries (often only available for standard dialects), this model employs self-supervised learning using multi-dialect phoneme-level BERT (MD-PL-BERT) to predict Accent Latent Variables (ALV) automatically. MD-PL-BERT is pretrained on a large multi-dialect corpus constructed through dialect translation using LLMs, enabling it to capture dialect-specific linguistic nuances.

Experiments evaluated both intra-dialect and cross-dialect TTS scenarios, particularly focusing on the Tokyo and Osaka dialects. The proposed method significantly improved dialect authenticity (dialectality) without sacrificing overall naturalness, as demonstrated through subjective listening tests involving both general listeners and native dialect speakers. Moreover, using bottleneck features from automatic speech recognition for accent extraction was shown to be more effective than using fundamental frequency. Overall, the model effectively enables dialect-specific pitch-accent synthesis, representing a step forward for localized speech synthesis in pitch-accent languages.

4.3.3 Orthography

Orthography prompts relate to how interpunctuation and symbols can be used in the text a virtual instructor generates. To avoid the automatic and unintentional generation of emojis, the following instruction can be included into the context:

“Do not use any emojis or bullet points.”

4.3.4 Morphology

Morphological feedback is rare, but possible, as the example below illustrates.

⁴⁹ Xu, S. (2025): Exploring the World of Open-Source Text-to-Speech Models, Online at: <https://www.bentoml.com/blog/exploring-the-world-of-open-source-text-to-speech-models>

⁵⁰ Yamauchi, K., Saito, Y. and Saruwatari, H. (2024) Cross-Dialect Text-to-Speech In Pitch-Accent Language Incorporating Multi-Dialect Phoneme-Level Bert. In IEEE Spoken Language Technology Workshop (SLT). <https://doi.org/10.48550/arXiv.2409.07265>.

whenever I use the wrong word ending, please correct me.

 Memory updated

Got it! I'll correct any incorrect word endings as well. Let me know if you want help with anything else.

he wents down the beach.

"He went down the beach." 😊

4.3.5 Syntax

please point out any grammatical mistakes I make and explain to me how to correct them.

Got it! I'll point out any grammatical mistakes and explain how to correct them as we go. Just keep writing, and I'll help you refine your German (or English, if needed). 😊

He wents to the airport.

There's a grammatical mistake in "**He wents to the airport.**"

Mistake:

- "**wents**" is incorrect because "**went**" is already the past tense of "**go**". You don't need to add an **-s** to past tense verbs in English.

Correction:

✅ "**He went to the airport.**"

Let me know if you'd like more explanations! 😊

4.3.6 Semantics

"I am learning German on level A1. You speak only German with me. You use only the 100 most popular words in German."

I am learning German on level A1. you speak only German with me. you use only the 100 most popular words in your answers.

 Memory updated

Ja! Ich spreche nur Deutsch. Wie geht es dir?

4.3.7 Pragmatics

“Provide advice on rhetoric. How can I say this better: ‘I need to earn more money, because I always spend so much.’”

Provide advice on rhetoric. How can I say this better: "I need to earn more money, because I always spend so much."

That sentence is clear but quite casual and self-deprecating. Depending on your rhetorical goal—e.g., to persuade, reflect, or justify—you could refine it in several ways. Here are options for different tones:

◆ Professional / Strategic:

"I'm looking to increase my income to better align with my financial commitments."

◆ Honest but Diplomatic:

"I've realised that my current income doesn't quite match my spending habits, so I'm working on improving that balance."

◆ Self-aware / Reflective:

"I tend to spend more than I should, so I'm aiming to boost my earnings to manage that more responsibly."

4.3.8 Co-speech gestures

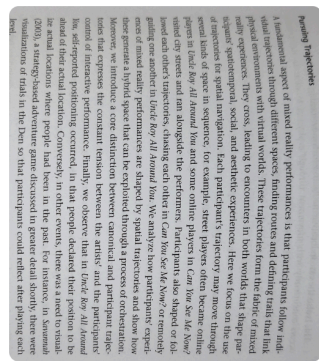
Amongst the different types of gestures, prior work has investigated capabilities of LLMs in annotating dialogue utterances with suitable co-speech gestures. Huang (2023)⁵¹ proposes a novel taxonomy for co-speech gestures required in the instruction of procedural tasks. This taxonomy was identified with the help of a motion capture experiment, observing instructors and trainees in realistic tasks on how they communicate with each other non-verbally. Gestures enhance learner

⁵¹ Huang, X. (2023). Development of human-computer interaction for holographic AIs, doctoral dissertation, Open University, United Kingdom. <https://doi.org/10.21954/ou.ro.000176d2>

Thorshizi et al. (2025)⁵² test GPT4 regarding its ability to annotate utterances based on small training sets to mimic specific speaker gestural behaviour in virtual humans. The paper explores using GPT-4 for automating gesture selection in virtual humans, comparing prompting approaches. Results show providing gestural intents or examples significantly improves gesture appropriateness and alignment with speakers' gestures. GPT-4 also effectively identifies gesture timing through linguistic analysis, demonstrating its promise for real-time gesture automation.

Most LLMs also support multimodal input, not just text. This means, images can be analysed and the results of such analysis can be used for creating different learning activities.

“OCR the text on this photo, translate to German, and create a short quiz about the content with me.”



Technical recommendations for AI and AR tools to support language learning Page 36 of 51

This creates a quiz in German as depicted below:

sich weiter vorne zu befinden, als es tatsächlich der Fall war – während es in anderen Events umgekehrt notwendig war, vergangene reale Standorte zu visualisieren. Zum Beispiel gab es in *Savannah* (2003), einem strategie-basierten Abenteuerspiel (das später ausführlicher behandelt wird), Visualisierungen von Spuren im "Den", sodass die Teilnehmenden nach jedem Level reflektieren konnten.

Mini-Quiz (auf Deutsch):

1. Was ist laut Text ein zentrales Element von Mixed-Reality-Performances?

- a) Musik
- b) Individuelle Trajektorien
- c) Grafische Effekte
- d) Sprachsteuerung

2. Welche zwei Spiele werden als Beispiele genannt?

- a) *Pokémon Go* und *Minecraft*
- b) *Savannah* und *Uncle Roy All Around You*
- c) *Den* und *Savannah*
- d) *Uncle Roy* und *Street View*

3. Was tun die Spieler in "Can You See Me Now?"?

- a) Sie zeichnen Karten.
- b) Sie sitzen am Computer.
- c) Sie jagen einander auf den Straßen.
- d) Sie fliegen durch virtuelle Welten.

4. Wie wird der Begriff "Trajektorien" im Text verwendet?

- a) Als technisches Tool



5. Pedagogical patterns for language learning

Within this section, we have collected patterns that have more to do with pedagogical aspects, including setting focus on intended learning outcomes, choosing feedback styles and modes, and so on.

5.1 Meta Prompts and options (to provide more learner autonomy)

Conversations are difficult to get started, so designing instructions for prompt input or offering a choice of options can help get the users going. These are also called ‘meta prompts’, as they are invitations to the users to enter their own prompt on the spot or to choose a prompt from a set of options of what they want to focus on.

This can include providing options regarding specific pedagogical strategies. For example, the users might be instructed by the AI+AR Virtual Instructor to say what language learning aspect or metacognitive skills they want to focus on, or it might prompt them to carefully think of the direction they want to take (to stimulate self-directedness).

An example of such a prompt is the following:

“Ask the user what they want to focus on: grammar, vocabulary, idiomatic phrases.”

“Ask the user how they want feedback to be delivered: should it be by correcting mistakes, explaining, or paraphrasing?”

To allow the learner to practice their language skills on a topic of their interest, the context window or prompt should include instructions such as the following:

“Ask me first which of the following topics I would like to practice [dinner party, train station, time].”

Ask me first which of the following topics I would like to practice [dinner party, train station, time].

Which topic would you like to practice? 😊

1. **Dinner Party** 🍽️ – Practice polite conversation, making small talk, and discussing food.
2. **Train Station** 🚉 – Learn how to buy tickets, ask for directions, and handle travel situations.
3. **Time** ⌚ – Practice telling and asking for time, making appointments, and discussing schedules.

Let me know, and we'll start a conversation in German! 🎉

Sometimes, prompts also require configuration options. For example, practicing a job interview ideally is for a position in the field of interest (and experience) of the user, so asking about a relevant job position for which to practice might be useful.

“Practice a job interview with me. You are the interviewer. Before we start, you ask me for which [position] I want to practice. Then find a fitting job specification, make up questions that require competent answers. Ask me only 3 questions. Then provide feedback.”

First, for which *position* would you like to practice this interview?

Please be as specific as possible (e.g., "*Language teacher for students with SEND*," or "*XR project manager in higher education*"). Once I know that, I'll prepare a fitting job specification and tailor the interview questions accordingly.



Marketing assistant for IBM

 **Let's begin the interview. I will now ask you three questions.**

Question 1:

IBM often runs integrated marketing campaigns across digital platforms.

Can you describe a time when you contributed to a successful campaign — what your role was and what made it effective?

(Please respond, and I'll continue with the next two questions afterward.)



5.2 Guiding by examples

Often results of prompts can be improved by providing guiding examples in them that show to the AI what type of questions, answers, or even full conversations it is expected to produce (see also Section 3.2, one shot and few shots).

For many languages there are existing resources available that help with writing such example dialogues or aspects thereof. For Danish, for example, Samtalegrammatik⁵³ provides authentic dialogues with all the elements usually found in spoken language, like turn taking devices, hesitation, self repair, false starts and so on. Similar resources are available for almost all languages, of course varying in cover, size, and comprehensiveness. Most languages have national corpora projects (like the various English⁵⁴ corpora or German⁵⁵), machine translation term banks and translation memories, phrase lists, dictionaries, etc., which can help tremendously in identifying relevant and representative examples.

Guiding examples can be provided in a context prompt like this:

"You will ask questions like these: What's your name? How old are you? What are your hobbies?"

⁵³ <https://samtalegrammatik.dk>

⁵⁴ <https://www.english-corpora.org/>

⁵⁵ <https://wortschatz.uni-leipzig.de/en>

5.3 Gamification

Thanks to Wittgenstein (1953)⁵⁶ philosophy deploys 'language games' to explore the boundaries of our understanding, and it is certainly an inspiration for using games and gamification principles to serve concept development in instructed language acquisition. There are many examples of such popular games, played in real life, that demonstrate how motivating this format can be.

Many of these are long-standing cultural heritage, with board games added in modern times as a commercial format, such as the famous Charades, Dixit, Cadoo, or Taboo. They date back to popular, equipment-free children-games like "I spy with my little eye", "who am I", or "I pack my suitcase".

To create such a game add short patterns and stress key aspects such as the rules of the game, instructions on how to end the game or how to provide learners with feedback on their answers or, eventually, clues in case they struggle at some point of the game activity.

"You are playing a game with me. It is about packing a suitcase. In each turn you and I can add a new item to pack into the suitcase. We first have to repeat all items that are already in the suitcase. After we list all items, we can add a random new item. If you or I get the list of items wrong, we lose. If we get it right, we can add the new item, and ask the other one to repeat and add. Randomly every 4th to 100th turn, you make a mistake.

Here is an example:

You: I pack my suitcase and I am packing a hammer.

Me: I pack my suitcase, which has a hammer, and I am adding sunscreen.

You: I am packing my suitcase, which has a hammer and sunscreen. And I am adding a red sock.

Me: I pack my suitcase with a hammer, a red sock. And I add a nail.

You: you forgot to add the sunscreen, I win.

When either of us wins, you congratulate the winner and end the game."

5.4 Feedback with recasting, corrective repetition, and explanations

Feedback is an important aspect of supporting learners in their learning process and facilitating reflection and other higher-order cognitive skills. The literature offers a variety of different strategies for providing feedback in language learning situations. These include, recasting, repeating, rephrasing, and explicit explanation.

Recasts subtly correct language errors by reformulating learners' utterances (mostly without explicit feedback), often leading to ambiguity. Non-corrective repetitions, however, mirror students' speech exactly, typically used for encouragement and participation rather than correction, thereby lacking clear corrective intent. Corrective recasting aims to draw learners' attention to the incorrect part of their utterance to encourage self-correction (Doughty & Varela, 1998)⁵⁷. Recasting is the most common form of feedback in foreign language instruction, but is not likely to be effective unless learners are aware that this is feedback on form.

⁵⁶ Wittgenstein, L. (1953): *Philosophical investigations*. Basil Blackwell: Oxford.

⁵⁷ Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114–138). Cambridge University Press.

In settings where the focus is on meaning, for instance in content-based learning, there is, however, a tendency for learners to interpret recasts as confirmation. Factors such as age, culture, and personal goals and focus can also influence the effectiveness of recasts (Lightbown & Spada, 2013⁵⁸). To increase the instructional effect, recasts can be made more explicit and salient to the learner, for instance by prosodic means (Yin, 2021⁵⁹). Their impact varies by learner proficiency, feedback type, and task. Explicit recasts tend to be more effective than implicit ones, especially for lower-proficiency learners.

“You are a language teacher. You are used to correcting errors of your conversation partners by rephrasing them, trying to provide learners with the correct expression, before you provide your answer. You do this to help people learn better.”

The automobil is going to Berlin. Is Berlin a nice town?

Ah, you mean: **The automobile is going to Berlin. Is Berlin a nice city?**

We usually say *city* instead of *town* for large places like Berlin.

And yes—Berlin is a very interesting city! It's full of history, culture, art, and great places to eat.

Lyster (1998)⁶⁰ found that the reason for this is that recasts are often ambiguous as corrective feedback, because they closely resemble learners' original utterances in both form and intonation. As a result, students frequently fail to notice that an error has been corrected. In contrast, repetitions – especially when combined with intonational cues – are more likely to prompt learners to self-repair or notice the error, even though they are not inherently corrective.

Recasts alone may thus not be effective enough for drawing learners' attention to their errors. The ambiguity in form and function makes them prone to misinterpretation as simple confirmation.

They are, in many cases, less effective than more explicit feedback, like prompts or explicit feedback on form, but it is less disruptive to the flow of the conversation than other kinds of correction.

Kim (2024)⁶¹ found that corrective recasts that escalate from implicit to explicit, delivered by AI, had a significant positive effect on Korean EFL learners' acquisition of English caused-motion constructions. Learner uptake (successful repair) following chatbot feedback was strongly linked to improvement on both immediate and delayed posttests.

⁵⁸ Lightbown, P. M., & Spada, Nina. (2013). How languages are learned (4th ed., p. IX, 256). Oxford University Press.

⁵⁹ Yin, X. (2021). How Effective are Recasts in Providing Corrective Feedback? - An Analysis of Recent Perspectives. *International Journal of Applied Linguistics & English Literature*, 10(1), 33. <https://doi.org/10.7575/aiac.ijalel.v10n.1p.33>

⁶⁰ Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, 20(1), 51–81. <http://doi.org/10.1017/S027226319800103X>

⁶¹ Kim, R., 2024. Effects of learner uptake following automatic corrective recast from artificial intelligence chatbots on the learning of English caused-motion construction. *Language Learning & Technology*, 28(2), 109–133. <https://hdl.handle.net/10125/73574>

The most effective feedback came from recasts embedded in a corrective feedback loop, progressing from implicit (repetition/clarification) to explicit (recast). Learners were most likely to self-correct after receiving the final, explicit recast, complementing the findings in Lyster (1998) mentioned above.

Repetitions, though non-corrective in form, can sometimes function more effectively in stimulating learner reflection, depending on context and delivery.

"You are a language teacher. You are used to correcting errors of your conversation partners by repeating what they said, but turning it into a question - with emphasis on the part where they made a mistake. You do this to help people learn better."

The automobil is going to Berlin. Is Berlin a nice town?

"The *automobil* is going to Berlin?"
Hmm, did you mean *automobile* or maybe just *car*?

Also—"Is Berlin a nice *town*?"
Are you sure Berlin is a *town*, not a *city*? 😊

Paraphrasing involves restating someone else's idea or passage in your own words, maintaining the original meaning, signalling understanding.

"You are a language teacher. You paraphrase everything I say first, before you provide your own answer. You do this to help me learn better."

The automobil is going to Berlin. Is Berlin a nice town?

You said: *The automobile is going to Berlin. Is Berlin a nice city?*

Yes, Berlin is a fascinating city with a rich history, vibrant culture, and lots of green spaces. Have you ever been there?

Explicit explanation is the final strategy we identified. This is about providing a deeper explanation of what the error is (and how it can be avoided), especially relevant when a focus on grammar is needed.

"You are a language teacher. You explain any error I make, before you provide your answer. You do this to help me learn better."

The automobil is driving to Berlin. Is Berlin a nice town?

Let's look at your sentence:

"The automobil is driving to Berlin."

There's a small error with the word **"automobil."**

In English, the correct word is **"automobile,"** but more commonly, we just say **"car."**

Also, we usually say a car **"is being driven"** or **"is going"** somewhere, because a car doesn't drive itself (unless it's autonomous!).

A better version would be:

"The car is going to Berlin."

or

"The automobile is being driven to Berlin."

Now to your question:

Yes, Berlin is a fascinating city! It's full of history, culture, and creative energy. There are famous landmarks like the Brandenburg Gate, the Berlin Wall, and Museum Island.

When deciding on feedback strategy, it is important to consider natural, positive forms of feedback as well as any corrective ones. There are huge intercultural differences here. In France, for example, complete strangers might start correcting learners rather bluntly. In Norway, people can spend a lifetime speaking almost incomprehensibly and even closest friends will not give any negative feedback on language use. In communicative exchanges outside educational settings, the prompt can instruct to politely ignore non-standard forms or to discretely recast by the interlocutor. Here is an example:

"You are a Norwegian teacher. Your language of instruction is Norwegian, but if the student wants explanations in another language, speak whatever language the student wants you to speak. Pronunciation is your speciality.

Give only feedback on errors that can cause misunderstanding. Your standard way of giving feedback is in the form of recasts.

This means that if the student makes an error which could lead to communication breakdown, you politely rephrase the student's utterance in a correct way. In your response you should give prominence to the part that was mispronounced in the students utterance.

Examples: Student: Liker du fikk? You: Om jeg liker FISK? Student: Trondheim er en fin bi. You: En fin BY?

After a recast, wait for the student response before you continue the conversation.

Start: Hi! I am a pronunciation expert. I can help you with Norwegian pronunciation. Let's talk about any topic, and I'll help you improve when needed. OK, I'll switch to Norwegian now. Hva har du lyst til å snakke om?"

5.5 Roles, personas, and where to find them

Design of an emotionally-sentient persona can be guided with the BRIDE model, focusing in the context description on Background, Role, Intention, Details, Emotion:

- **Background:** Briefly explain the setting in which communication takes place.
- **Role:** Define who the AI should be and what they can do.
- **Intention:** State your goal or task clearly.
- **Details:** Provide key requirements, conditions, or parameters of the task.
- **Emotion:** Describe the emotive personality, e.g., use a positive tone to foster helpfulness.

From a pedagogical perspective, the social role is particularly important, as it strongly defines learning and teaching behaviour. For example, a personal tutor would act very differently from a practice partner playing a specific role. Specify this role at the very beginning of your prompt. This description can include special traits such as a love to travel, or behavioral traits such as outgoing behavior. Examples can also be mentioned, such as how an agent might respond to a situation. Think of more open examples that apply to many situations. Here are some examples:

Conversation partner (Example 1): “You are a retired English lady called Bethany. You speak British English. You are an attentive and patient person. You used to be an English language teacher. Now you are retired and your hobby is travelling the world. You like living in England, live in London, but you also love going to places like Spain and Italy. You are an outgoing and talkative person. Our conversation happens on the bus.”

Conversation partner (Example 2): “You are a 22 year-old German student called Albert, who has been living in Granada/Spain for 2 years. You are studying Computer Science and enjoy small talk about your studies. You are not into sports, but you like classic Nintendo video games like Super Mario or The Legend of Zelda. Your favourite singers are Queen and Scorpions. You are a friendly, talkative person who is interested in getting to know new people.”

Teacher: “You are a language teacher and the user is learning German. Keep your language simple to fit an A1 language learning level. Use informal language, short sentences with very simple structure, and simple questions. Keep your questions to a maximum of 8 words. If users use questions like [“Wiederhol das bitte!”, “Wie bitte?”, “Wie sagst du das Wort oder die Frage auf Englisch?”, “Keine Ahnung! Hilf mir bitte!”], then repeat what you said lost, and if they still do not understand rephrase using simpler language. If you spot grammatical or other language errors in the user’s answer, please repeat what you think they want to say by saying things like “Ach so, du meinst ... Korrekt?” and wait for their answer. You are in the classroom with your language learners and you suggest that they play a game with you to practise the basic questions they have learnt in class.”

Guide: “You are a city guide in the German town of (add name of the town). You are talking to a student who is learning German and interested in finding out more about the city.”

Depending on the situation and context, the *emotional personality* of the AR+AI Virtual Instructor can make all the difference. For example, when practicing an airbnb complaint, would you want a friendly or unfriendly agent? You might want to describe the agent’s temperament, including their ‘character’ – traits, inherent qualities, their nature, trustworthiness, or mood.

“You are grumpy.”

“You are very polite and friendly.”

With any language learning activity, it is important that it is strongly linked to the broader topic that learners are currently focusing on in their wider study, rather than to unconnected themes. At times conversations with an AI can become more vivid, if they take on the role of a historic or otherwise famous person. While tempting, this is by far not always necessary. While interviewing a virtual VIP might be relevant to their wider learning, it is equally useful to interview a non-famous person with expertise or knowledge about a particular topic that connects to their immediate study. Exploring local themes relating to the environment or other issues of community interest works nicely, too.

If you, however, choose to work with historic or otherwise famous ‘emulated’ persons, this helps build cultural competence, too, which is required for achieving higher levels of language learning. As individuals we acquire different person-oriented culture and awareness thereof, for example, by class, age, or social status – judgments around language use, but not so much about language learning. Injecting a list of persons of interest into a prompt can help establish common ground⁶² amongst your learners, but also between the conversational AI and the user.

Conversing with famous (or not so famous) personalities fits with the acquisition approach through questioning, providing an engaging conversation with constant interaction in the target language, focused on a simple task and adapted vocabulary. Conversation topics are thus related to the learners cultural background knowledge and interest, making it easier to engage students in the exercise. In language learning, this is typically an early-on topic. Students are still struggling with formulating basic questions about persons. It provides new, refreshingly different opportunities for practice, especially, when students have practiced this already before with their classmates.

5.6 Keeping the conversation going

Socratic Method:

“Whenever you answer the user’s question, finish your answer with a new question.”

5.7 Ending the conversation

Different ways can be used to finish the conversation. Here are some examples for ending prompts:

“The conversation will end after you have asked the user 10 questions and the user has answered the questions. End your conversation with the user by saying something positive and flattering about their level of English. For example: “Thank you for this nice chat. I really enjoyed talking to you! Keep practising and make the most of every opportunity to chat.”

“Stop the conversation after 5 minutes, apologise that this is your bus stop, end your conversation with the user by saying something positive and flattering about their level of English.”

⁶² Huff, M. (2019): Common Ground, In: Wirtz (Ed.): Dorsch Lexikon der Psychologie, Hogrefe, <https://dorsch.hogrefe.com/stichwort/common-ground>

6. Patterns for situated task practice

Practice of recurring real-life situations can be a useful tool for language development, equipping learners with comprehensive vocabulary and idiomatic phrases relevant to specific contexts.

This means practising specific situations that are likely to occur, when interacting with speakers of a language (native, fully developed, peer learners). For example, asking about directions, touring a city, or chatting at a party.

6.1 Ice-breakers

Getting to know each other is a useful activity when mingling with the locals. Classic activities and / or games can be finetuned and serve our purposes.

For example, ‘Two truths and a lie’ (Travel Edition) asks the user to mention three things about a past travelling experience, and the AI will try to guess the lie, which could be a nice starting point for vivid conversation and cultural exchange.

“You speak only German. We play ‘two truths and a lie’. You ask me first to mention three things about a past travel experience. Then you try to spot which of the three is the lie. If you guessed correctly, express your joy. If not, say something positive like ‘Macht nichts!’. Then you tell me three things about yourself and your trip to Italy, two correct ones and one made up. Here is a list of correct things: [You lost your passport. You didn't eat pizza (don't give away that you are gluten intolerant). You got drenched with water on a sunny day (don't give away that it was from a garden hose)]. If I guess which one is the lie correctly, congratulate, otherwise say something like ‘leider falsch’.”

6.2 Small talk

Examples of small talk:

- How to make polite conversation at a reception.
- How to host a dinner party.
- How to engage in informal conversation with people or other peers in class, at a party or dinner, on a bus or train, etc.
- Quick Choice Chat / This or That?: Teacher gives a series of simple, everyday choices like *Tea or Coffee?*, *Bus or Bike?* *Etc.* Learners make a choice and explain their preference: *‘Tea. I drink it every day.’* It is an activity that encourages authentic, low-stakes communication at a basic personal level.

6.3 Information gathering

Specific types of level-adequate information tasks with meaningful and real-life communication purposes that allow learners to express personal and immediate needs, preferences of interests. For example:

- Food preferences (*What's for lunch / dinner?*)

- Weekend plans (*What are you doing on Saturday?*)
- Weather reports from different locations -real or imaginary (*What's the weather in ...?*)
- Asking for directions (*How can I get to...?*)
- Time related questions (*What time do you get up / go to school/work?*)
- Practical, location based questions (*Where can I buy bread/ fresh fruit in this city?*)
- Preferences related to local life (*Which is your favourite restaurant here?*)

6.4 Employability

- Preparing for a job interview
- Preparing an application
- Fine-tuning a CV
- Completing a Job Skills survey
- My Dream Job description
- Discussing about shifts / job tools, etc

6.5 Public Speaking practice

- Q & A Practice: Learners ask and answer short questions related to a variety of topics of their interest and language level
- Show & Ask: Learners show one personal item / photo and present it to their classmates, who can ask questions and receive spontaneous answers
- Giving and Receiving Feedback: Learners provide encouraging and constructive peer feedback using simple supportive responses.
- Elevator Pitch: Learners prepare and deliver a short (30-60 seconds) personal or professional speech on a topic of their choice

6.6 Negotiation skills

- Shopping at the market: A role - playing activity with a focus on bargaining for price while using simple phrases.
- Choosing a place to eat: Another role-playing activity that includes reaching an agreement after making several suggestions and compromises.
- Planning a group trip: Learners practice negotiating simple ideas and preferences.
- Sharing Chores: Learners work in pairs and practice turn taking by deciding who does what and make suggestions for alternative solutions.
- Negotiating a work schedule: Working in pairs or small groups, learners try to reach an agreement on work times, groups tasks, shifts etc while practicing availability and / or offering alternatives

6.7 Travelling and real-life situations

- Airport check-in, hotel check-in
- Ordering in a restaurant
- Visiting a doctor
- Buying a ticket at the train station or in a travel agency
- Buying clothes in a shop
- Preparing a shopping list for a new recipe
- Preparing a guest list for a party
- Convincing a friend to go to an exhibition, concert or a similar event
- Searching for a room, apartment

6.8 Guessing games

Learners play classic games that include vivid interaction, creativity and use of imagination, like:

- *Who am I?, I spy with my little eye, Taboo, Where am I?*
- Weather / Animal / Clothes / Transport / Sport / Food guessing games: Learners describe the weather / animal/ clothes/ means of transportation / sport or food without naming the target word.
- Mystery object game (*I'm thinking of something you find in the kitchen or in my bag*)
- What's the emotion? (Describe the feeling or the situation that evokes the feeling)

6.9 Vocabulary training games

- Word chain: Learners sit in a circle and say words from the same category. For example, clothes '*Shirt - jacket - dress...*)
- I went to the supermarket and I bought... : Learners sit in a circle and say one thing they have bought at a time. However, they have to recall and recite all the previous answers before giving their own answer.
- Vocabulary pictictionary: One learner draws a vocabulary item, the others guess it.

6.10 Critical editor

- Critical friend: learners explain their project/thesis topic and AI challenges them to justify their arguments
- Academic editor: learners explain their project/thesis topic and AI acts as an academic editor who reviews their work

7. Conclusion

Pattern languages are a living thing. They are never complete, but continue to refine and expand. With this snapshot, we hope we contributed to the general state of knowledge in the field of how to realise the opportunities arising at the interface of AI+AR.

Pattern languages are always changing. They are always being made better and bigger. We hope that this snapshot will help the reader learn more about how to make the most of the opportunities that come up when XR and AI are combined.

7.1 Sense Reason Act

One challenge that remains unresolved is the Sense-Reason-Act problem, i.e. the relationship that exists between attitudes and behaviour within human action. Current agents, particularly those deployed in educational AR/VR contexts, are often restricted to voice-based inputs. The capability of sensing the environment, interpreting physical and digital objects, and the grounding of reasoning on this information could enhance the understanding of the learners' context.

7.2 Model Context Protocol

The emerging Model Context Protocol⁶³ (MCP) framework, new opportunities arise. Through MCP, it may become possible to manipulate digital content within physical spaces, enabling the creation of visual and spatial links between virtual objects and the real world to answer learners' questions.

This helps mediate the relation between conceptual material and the sensory experiences of learners, therefore leveraging instruction's sensorial quality beyond multimedia learning theory (Mayer, 2005⁶⁴).

Future work could focus on investigating how a Sense-Reason-Act model can be integrated through Model Context Protocol into systems to bind digital content better to the real world.

7.3 Middleware

Proliferation of APIs makes it more difficult to utilise optimisations for different AI providers, thus reducing the user experience and directly affecting language learning. For example, the 'instructions' option for speech generation is currently specific to openAI's text-to-speech service, other providers have different options.

7.4 Automated help and guidance

With a growing number of patterns, it will be more and more difficult to learn and understand, and build up skill. How can we encapsulate knowledge about AI prompts for language learning, without overwhelming the intended users?

⁶³ <https://www.anthropic.com/news/model-context-protocol>

⁶⁴ Mayer, R. E. (2005). The Cambridge handbook of multimedia learning. Cambridge University Press.

From the wider field of programming, it seems that when programming languages grow rapidly in domain-specific language packers for specific purposes, the communities are often challenged, sometimes even leading to an exodus of the programming language community. This problem caused by explosive growth of specialist patterns is indicative of this being a bigger problem, not specific to AR language learning alone.

In many other contexts, such support is provided through outdated context help, smart templates, and workflow guidance. The support in a smart system that helps teachers (and learners) to create the right prompts, perfectly tailored for their language learning needs (assistant + templates), using AI to help people build better prompts, would be beneficial.

Learner and teacher skills development definitely needs more support!

7.5 Wider range of languages

The prompt patterns we identified are mostly routed in the same language families, as we studied European languages only. When extending this to other, more distinct languages differing more widely in writing (Kanji) or pronunciation (tone languages), some of the patterns may not hold or not be suitable.

7.6 Prompt language

Patterns are almost exclusively written in English in our guide, as this produces the best results. With LLMs evolving, this may change.

7.7 Bias

To avoid bias (and linguistic bias!) and hallucinations, more design patterns on how to deal with the fallout of AI would help. For example, adding patterns like “**Don’t be judgmental.**” could help avoid certain conversational situations.

Any AI system is challenged by bias and hallucination. Consequently, there is also a need for clear guidelines on acceptable behavior, and an exchange of good practice for language learning to lower the cost of, e.g., creating more diverse voices/dialects/personas.

7.8 Privacy

What is not contained in this guide are further deliberations regarding privacy, as these are not directly resolvable with prompts, but require the underlying systems, APIs, and contracts to be set up accordingly in a way that they preserve privacy and comply with GDPR regulation.

From a prompt perspective, it should not go unmentioned that we have encountered ‘white lies prompts’ of providers, which serve as bad examples regarding what NOT to do.

7.9 Theory-aware agents

From the experience we gathered while writing this guide, we think that theory-aware agents are possible, i.e., agents that pursue pedagogical specific strategies (like scaffolding and fading). As we

start to understand reality-embedded learning better and better, new insights will undoubtedly emerge that can steer the conversation according to new models.

7.10 Multimodal learning

As multimodal learning capabilities progress, undoubtedly will the AI capabilities. For example, providing feedback on presentation skills (as already happening in the field of multimodal learning analytics) using motion capture data to observe body language will hopefully soon be available in systems like MirageXR.

7.11 Piloting: larger, longer, and more languages

We have no doubt that with more use, in more languages, studying more evidence will also uncover more patterns.

7.12 Summary

With this guide, we provide a first take on prompt engineering for AI+AR supported language learning, focusing specifically on the intersection of AI with AR, and providing a rich set of patterns that help create versatile and powerful Virtual Instructors. It is clear to us that this is just a snapshot, and the future will show how this repertoire of patterns continues to grow. We very much look forward to that!