Designing an LLM-based Assistant for Speech Therapists: the Case of e-SpeechT

Miriana Calvano[1], Antonio Curci[1,2], Antonio Piccinno[1] and Alfonso Pio Pretorino[1]

[1]*Department of Computer Science, University of Bari Aldo Moro, Via E. Orabona 4, 70125, Bari, Italy*
[2]*Department of Computer Science, University of Pisa, Largo B. Pontecorvo 3, 56197, Pisa, Italy*

Abstract

The Large Language Models (LLMs) are increasingly being employed across many fields, such as healthcare, finance, and automotive, revolutionizing how humans carry out activities. In this research, the integration of an LLM-based assistant in *e-SpeechT*, a web application that supports speech therapists in conducting, administering, and monitoring therapies, is investigated. It is intended to integrate an LLM component that allows the therapist to be supported in the generation of exercises to administer during the therapy, establishing a symbiotic relationship between the two parties. This approach reflects the core ideas of End-User Development (EUD) by enabling therapists to directly influence, customize and manage system behaviour without relying on advanced programming knowledge. The system enables the therapist to share the created exercise, reinforcing the concept of Cultures of Participation (CoP). Using Prompt Engineering principles and techniques, many prompts have been defined and underwent a multi-stage evaluation pipeline to assess the quality of the generated outputs. The obtained results show that a well-structured and clear prompt can generate consistent and reliable outputs that align with the prompt instructions.

Keywords

Large Language Models (LLMs), Prompt Engineering, Symbiotic Artificial Intelligence (SAI), Speech Therapy

1. Introduction

Speech Therapy is a discipline that focuses on the study and treatment of disorders affecting language, communication, and cognitive disorders [1]. They affect both children and adults, manifesting in a variety of forms whose triggering factors are often challenging to identify [2, 3]. The speech therapist's role is to address these disorders, leveraging re-education methods and personalized therapies. During the treatment, the speech therapist must employ the most suitable methods and approach for each individual case, taking into account the specific clinical status [4, 5].

Integrating Artificial Intelligence (AI) in healthcare has recently brought many benefits, such as automatic transcription of medical reports and personalization of treatments reducing the amount of time and human resources to carry on a therapy. Russell and Norvig define AI as "The study of agents that receive percepts from the environment and perform actions" [6] and, among the different AI techniques, LLMs have gained great attention in the past few years. Relevant examples include OpenAI's series, Google's Gemini, and Anthropic's models. They have demonstrated state-of-the-art performance on various Natural Language Processing (NLP) benchmarks, demonstrating how LLMs can impact both academia and industry [7, 8, 9]. LLMs, such as GPT and LLama, have further accelerated innovation, allowing developers to automate generative tasks such as report generation and conversational agents [10, 11].

This research investigates the integration of an LLM-based assistant in an already-existing system, called "e-SpeechT", to support speech therapists in their work by generating therapeutic material (i.e., exercises). During the creation process, the therapist can decide to share the exercise with all the

Proceedings of the 10th International Workshop on Cultures of Participation in the Digital Age (CoPDA 2026): Exploring the Relationship between EUD, AI-Assisted Development, and Meta-Design, June 2026, Venice, Italy.
miriana.calvano@uniba.it (M. Calvano); antonio.curci@uniba.it (A. Curci); antonio.piccinno@uniba.it (A. Piccinno); a.pretorino@phd.uniba.it (A. P. Pretorino)
0000-0002-9507-9940 (M. Calvano); 0000-0001-6863-872X (A. Curci); 0000-0003-1561-7073 (A. Piccinno); 0009-0000-8422-7808 (A. P. Pretorino)

therapists registered on eSpeechT or keep it private. This can improve their knowledge and ability to adapt to new situations, prompting CoP, in which users actively collaborate, exchange knowledge, and co-create resources [12][13]. The active role of the therapists is reinforced during the exercise-creation process, reflecting the pillars of CoP, which is crucial in this research. In this phase, they can choose to make the exercise private, accessible only to the creator, or public, allowing it to be viewed by any therapist on the platform. The ability to share exercises promotes professional exchange and reuse, encourages collective reflection on them, and ultimately strengthens the sense of belonging within a CoP. However, this integration introduces several constraints that must be considered because in high-risk domains, such as healthcare and medicine, it is essential that AI-based components are created following a human-centred approach, ensuring that humans are supported rather than replaced. This relationship is embodied in Symbiotic Artificial Intelligence (SAI), a specialization of Human-Centred Artificial Intelligence (HCAI), which emphasizes continuous collaboration between humans and AI [14]. The LLM-based component presented in this research was designed and integrated into e-SpeechT following a "symbiosis-by-design-approach". In this perspective, SAI ensures that humans remain *in-the-loop* throughout the decision-making process, actively supervising and guiding the LLM at each stage. This approach aligns with the pillars of EUD, which empower domain experts to actively shape, adapt, and control system behaviour without requiring advanced programming skills [15, 16, 17]. By integrating EUD practices within a SAI approach, speech therapists are not only passive users but they actively modify and shape the system, being capable of refining outputs and steering system behavior based on their expertise. Such functionalities have been integrated into e-SpeechT with respect to the creation process of different exercises.

The article is structured as follows: Section 2 presents some applications of AI in speech therapy together with a description of e-SpeechT; Section 3 explains the methodology of the conducted experiment; Section 4 presents the generated exercises; Section 5 draws the conclusions and outlines directions for future work.

2. Grounding e-SpeechT in the Literature

As AI is providing substantial advantages to the field of healthcare, it is also bringing limitations and gaps that must be addressed: while the technological potential can be impactful, real-world adoption depends not only on the model used but also on robust validation protocols to ensure a responsible interaction process. This section also explores these topics and reports a description of e-SpeechT, highlighting its main actors and its functionalities.

2.1. Related Work

The use of AI in speech therapy is still an emerging field, but it reflects broader trends in healthcare innovation. Recent studies and surveys reveal that many researchers are exploring AI-based tools with cautious interest. Initial integrations of AI in healthcare have focused on managerial support, such as generating patient reports or streamlining documentation [18, 19].

Current literature lacks fully-implemented solutions in which LLMs directly supports therapists by creating customized exercises or interactive materials. Nevertheless, some academics investigate the role of LLM in speech therapy. For instance, Du and Juefei-Xu evaluated the potential of ChatGPT for clinical utility. With the help of an American Speech-Language-Hearing Association (ASHA)-certified speech-language therapist [21], they assessed overall performance on various speech therapy tasks by evaluating ChatGPT's ability to generate therapy materials in the ASHA "Big 9" domains (e.g., articulation, cognitive aspects). While the therapist's assessment was mostly positive, the authors stressed the need for the functionality that allows professionals to review to correct occasional inaccuracies [20]. Similarly, Cho et al. investigate the effectiveness of LLMs in speech therapy with high-functioning autistic adolescents. The study used an LLM assigned a specific persona, shaped by a set of instructions aided the LLM to adopt an empathic and appropriate communication style. The author aims to develop an interaction paradigm that is close to real-life situations allowing to conduct a complete assessment. The model was evaluated by experts in the field through a scorecard developed by the authors. It emerged that the LLM can offer benefits for adolescents with high-functioning autism [22]. Berrezueta-Guzman et al. integrated ChatGPT-4o in a robotic assistant for Attention-Deficit/Hyperactivity Disorder (ADHD) therapy. The models enable the robot to interact with the patient in real time, rather than serving as a static therapy tool. It is able to adapt and deliver personalized therapeutic support, also based on the robot's emotion-recognition capabilities. By considering this additional data, the LLM can adapt to any situation, keeping an appropriate communication style. Results show that integrating ChatGPT-4o into the robotic assistant has great potential for ADHD therapy. Thus, authors underline that it is essential to conduct real-world validation to unlock the full potential [23].

2.2. Overview of e-SpeechT

e-SpeechT is a web-based platform that supports speech therapists in diagnosing patients and creating and managing therapies [24]. The target audience of the platform involves three actors, each playing a meaningful role and entitled to different functionalities: speech therapist, caregiver, and patient. The *speech therapist* can manage patients, assign a set of exercises, review the status of ongoing therapies, and view the status of each assigned treatment. There are three categories of exercises: (1) Image Denomination exercise, which is composed of an object and a list of hints (at most 3) and the patient must recognize the depicted object and correctly name it; (2) Word Repetition, which has three different words that the children need to repeat; (3) Minimal Pair composed of a pair of words that differ in only one phonological element, thus having two very distinct meanings. The *Patient* is a child aged between 4 and 8 years old; they carry out the exercises assigned by the speech therapist, accessing their personal page by authenticating themselves through a personalized graphical password. The *Caregiver* monitors the overall status of the treatment of the patient they are responsible for and manages the child's personal profile (e.g., changes the password and appearance of the system) [25][26].

3. Integrating LLM in e-SpeechT

After conducting an interview with a speech therapist who is already familiar with e-SpeechT, the need for additional an AI-based assistance in creating therapeutic materials emerged. This need translates into an assistant that can speed up the creation of exercises while ensuring they are as diverse as possible compared to existing ones. Thus, an LLM-based component is designed to support the speech therapist in creating personalized exercises for patients. Following the EUD approach, the new functionalities aim to enable the speech therapist to interact with the LLM through mechanisms that ensure direct manipulation of the outputs and customization. For example, through structured and dynamic interfaces that guide therapists in modifying these artifacts (e.g., requests and outputs) based on their expertise. The interaction mechanisms, and consequently the user interfaces, are designed using a "symbiotic-by-design" approach that complies with the principles proposed by Calvano et al. [14].

3.1. Methodology

The LLM-based component was created employing the principles of SAI (i.e., Automation Level, Fairness, Protection, and Fairness) [14], ensuring that proper human-AI collaboration mechanisms are put in place. Specifically, allowing the speech therapist to access the "Assisted Creation" functionality, enabling the clinician to interact with the LLM actively, by remaining *in-the-loop* during the entire exercise creation process. The *Protection* principle imposes constraints on the model's technical deployment. The LLM must run locally (on the same machine that's running the system) rather than use remote services (e.g., APIs), because the prompt may contain patients' confidential data.

The experiments were conducted on a standard personal computer with average computational performance. This has been intentionally done to reflect the server's specification for where both the model and system will run. The latter influenced the choice of the LLMs to select and test. Five of

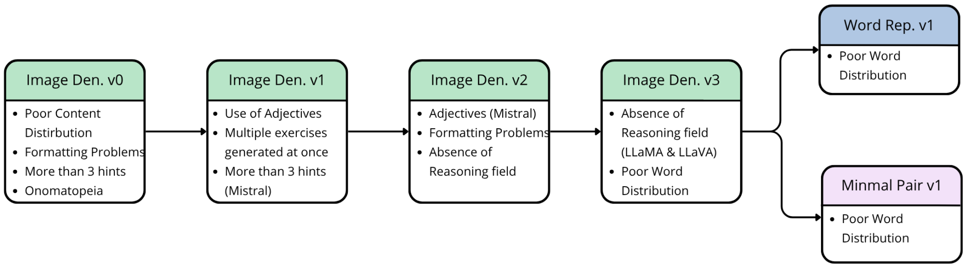

Figure 1: Prompt Versioning with its related problems

them were chosen: LLaVA:7b, Gemma3:12b, Llama3.1:8b, Mistral:7b, and Phi4:14b. An additional model, LLaMANtino-3-ANITA, was chosen for carrying an extra and final experimentation, focusing on the generation of exercise in Italian. These models are not fine-tuned on the exercises that are intended to be created with LLM. Thus, an extensive phase of Prompt Engineering will follow to assess the optimal structure and parameters for generating the exercises.

3.2. Prompt Definition and Refinement

To instruct the models to generate the three exercise types, the best practices and techniques of prompt engineering were employed; this discipline was defined by James Phoenix and it focuses on the design, refinement, and optimization of prompts to effectively guide the outputs of LLMs and other Generative Artificial Intelligence systems [27]. The techniques that employed are: Role Prompting, in which a specific sentence that assigns a precise role to the LLM is written into the prompt (in this study "You are a speech therapist [...]") [28]; Shot Prompting "family" (including zero-, single-, few shot prompting), where examples of the final output are inserted into the prompt, respectively zero, one or more than one examples [29, 30]. In addition, the experiments were run with different *temperatures*, which indicates the level of randomness in the model's output. Finally, a script for automatic prompting is created to facilitate the generation and collection of the generated exercises. It iterated over the five selected models, the temperature (from 0 to 1 with a step size of 0.1), role mode, and shot mode. In total, 330 outputs (66 per model) were obtained and stored in a file for further analysis and evaluation.

The generated exercises were evaluated through an ad hoc prompt evaluation pipeline involving three researchers; this step was necessary as it allowed to assess and address potential issues arising from poor LLMs performances or unclear specifications in the prompt. In the first iteration, the outputs were reviewed by a single researcher: if issues were identified and could be addressed through prompt revision (e.g., presence of special characters), the prompt was adjusted, and the generation process was repeated. Once the initial researcher determined that most of the generated exercises complied with the instructions specified in the prompt, it was considered *"Stable"*. At this stage, the second iteration was carried out, and the two additional researchers proceeded with a further evaluation of the outputs. A prompt was labeled as *Stable* when it is the final version for an exercise type, and it could be seen as the *stop criterion* for our iterative refinement pipeline. There are two specific conditions in which the prompt could be defined as such: the former is when the prompt has reached an acceptable level of reliability (most of the exercises do not have major issues, e.g., hallucination); the latter is when the problems are not fixable by adjusting the prompt, so a new approach is required to fix them (e.g., the unconditional repetition of the same words).

The first exercise type addressed is the Image Denomination exercise, whose prompt underwent many refinement iterations before being considered *Stable*. The third and final version (Image Denomination v3) serves as the foundation for developing the prompts used to generate the remaining exercise types. The diagram shown in Figure 1 lists the prompts, their respective versions, and the problems affecting them.

The third version of the prompt (v3) solved most of the issues that were encountered during the testing phases, and the integral version is reported in Section 3.2. The following major refinements were applied, leading to the final version:

- In the prompt, tags were added: *role* and *examples* tags. The former is used to add the role for the Role-Prompting technique; The latter is used for the zero-, single-, and few-shop prompting. They will be removed or substituted accordingly through the script used to run the experiments.
- The prompt was written with a specific structure, listing: instructions, context, input data, and output indicator [31].
- It's important to highlight the presence of precise instructions (rows 7 and 9). They have been shown to have a significant impact on the choice of objects, enabling the model to use more unique words compared to the previous prompt versions.

Image Denomination Prompt v3

```
1  {role}
2  Create an exercise for young children with age between 4 and 8. The exercise is called ”denominating images” and it is
       composed of an object, hints, and rationale.
3
4  Before choosing the object:
5   - Randomly select a category from the following list:
6     furniture, kitchenware, clothing, tools, body parts, nature, musical instruments, school supplies.
7   - Based on the selected category, choose a real, simple, and recognizable object that a child could identify.
8   - Avoid using the following overused objects: car, dog, cat, apple, banana, ball.
9   - Do not choose an object that is among the top 10 most common in early childhood vocabulary or learning materials.
10  - Do not use adjectives when naming or describing the object.
11
12 The hints must:
13 - Include 0 to 3 hints
14 - Be simple, short, and clear
15 - Describe physical features or typical use (e.g., you can ride it, you use it to write)
16 - Optionally describe features of the word (e.g., It has 5 letters, It starts with Pe___)
17 - Do not include sounds (e.g., It meows, It goes vroom!)
18
19 The rationale must include:
20 - The reasoning behind the object selection
21 - Why the object supports linguistic, cognitive, social, or motor development
22 - Why the object is appropriate for this exercise
23
24 Use the following format exactly:
25 Object: <object_name>
26 Hints
27 - <hint 1>
28 - <hint 2>
29 - <hint 3>
30 Reasons:
31 <your reasoning>
32
33 {examples}
34
35 Requirements (mandatory):
36 - Use plain text. Do not use special formatting like bold text (**), italics or underlined text.
37 - Create only one exercise
38 - You must select a simple subject appropriate for a child
39 - Do not generate the image
40 - Do not use any of the example objects listed above
41 - Just write the object and hints
42 - Do not include any additional information outside the required format
```

4. Results

The multiple refinement steps were necessary to address relevant issues that were affecting the LLMs output. Afterwards, each prompt was tailored to the exercise type, and the final output, in most cases, aligned with the given instructions in the prompt, without any formatting mistakes or hallucinations. Table 1 shows one good example for each exercise type, generated by the different models with the corresponding *Stable* prompt. The generated exercises were similar to those created by speech therapists in both form and style. For the Image Denomination exercise, a minor issue that still affects the output generated with the last version of the prompt is poor word distribution (frequency of unique words). This issue was observed since the early prompt versions, and various attempts were made to improve the distribution. By taking Phi4:14b model as reference, in *Image Denomination v2*, the number of unique words used in the exercises was only eight, while in the *Image Denomination v3*, the words used were eighteen, which is $\approx$ 220% more in the total count with respect to v2. The other models had a smaller increase in their use of unique words, on average $\approx$ 10%. To solve these issues, it was determined that a different approach is needed.

Another problem affecting this type of exercise was the absence of explainability, often called "Reasoning" in the context of LLMs. Together with the Objects and Hints, the prompt asks the model to explain why these objects and hints are selected. This issue affected the output in version v0; it was resolved in v1, but it reappeared in the last two versions, limited to LLaMA and LLaVA models. Twenty-nine out of one hundred and thirty-two have no reasoning field. Since the problem affects only two of five models, both with a low number of unique words, and there is a model that is performing much better than them, the issue was not addressed.

Image Denomination Exercise	**Word Repetition Exercise**	**Minimal Pair Exercise**
Object: - Tree Hints: 1. It begins with "Tr" 2. It has leaves 3. It has a trunk	Words: - Table - Pencil - Leaf	Words: - Sack - Back
Reason: The spoon was selected from the kitchenware category [...]	Reason: This exercise is designed to enhance linguistic [...]	Reason: The words sack and back form a minimal pair that contrasts [...]

Table 1
Examples of Generated Exercises

For the Word Repetition exercises, the model with the best performance was Phi4:14b. Among the five different models, it is the one that used the most variety in the words and it's the most reliable model, since almost all of the exercises generated by this model do not have formatting or hallucination issues. It was determined that such issues are not fixable through prompt refinements, and for this reason, the experimentation continued with the last exercise type.

For the Minimal Pair exercise, the best model is Mistral:7b. The main issue that was affecting the other four models was the use of the same pair across many exercises. For instance, Phi4:14b was the worst model for this type of exercise; of the 66 generated exercises, 34 of them had the Minimal Pair *"Cat, Hat"*, and most of the remaining ones had the word *"Hat"* in the pair. For this final type, further refinement was not feasible, since similar behaviour was observed in the other two exercise types.

Given the quality of the generated exercises, an additional experiment with LLaMANtino-3-ANITA was conducted: it is an instruction-tuned version of Llama-3-8b-instruct [32]. The final prompts for each exercise type were translated into Italian, and the same evaluation pipeline (used for the English prompt) was applied to the Italian prompts. The generated exercises, like the English ones, respect both the style and format details in the prompt. Two major problems affected them: the first concerns the absence of reasoning in the generated output (as observed with LLaVA and Llama), in which fourteen out of one hundred ninety-eight exercises lacked reasoning. The former regarded the type of words used in the exercise. The model generated outputs that used words like "Fucile" (in English "Rifle"), and the use of such words in this sensitive and high-risk domain must be avoided. A potential solution that may solve the problems encountered across the different experiments could be achieved by implementing in the prompt a specific "Aware" component, which is discussed in the next section (Section 5).

Integrating the LLM in e-SpeechT, the system enables therapists to intervene in the exercise-creation process by modifying it (either manually or with the support of an LLM) when necessary. This design reflects the core principles of EUD, as it allows therapists to influence, customise, and manage the system's behavior.

5. Conclusions

This research presents the results and a preliminary evaluation of an LLM-based component to integrate in e-SpeechT, a web application to support speech therapy. The generated exercises were evaluated using a multi-stage evaluation methodology involving three researchers who intervened at two separate moments. To instruct the model to generate the three exercise types, the prompt structure was designed in light of techniques and principles belonging to prompt engineering. The findings suggest that through a well-structured and clear prompt, it is possible to generate reliable exercises that meet the requirements specified, which compensates for the lack of transparency and non-determinism in these LLMs. As it stands, the prompts can be viewed as a "System Prompt" that provides the model with instructions, constraints, and goals for producing the output, which must always be revised and confirmed by the Speech Therapist before creation.

Considering that the therapist must give additional insights concerning the LLM-generated exercise, it is planned to investigate the integration of two independent components in the prompt: the first is the "Dynamic Prompt Component", in which the Speech Therapist provides additional information, enabling the specific exercise to be tailored to the clinical case. The second, named "Prompt Aware Component", will enable the system to firstly, track the words used in the exercises by directly accessing the database via Model Context Protocol (MCP) [1], secondly to avoid the use of "dangerous" words, if a generated exercise contains one of such words, the Speech Therapist will be able to forbid them. This could guarantee a more balanced word distribution across the different exercises and control over the words used, ultimately driving the LLM towards context-appropriate word selection.

Acknowledgments

The research of Miriana Calvano, Antonio Curci, and Antonio Piccinno is supported by the co-funding of the European Union - Next Generation EU: NRRP Initiative, Mission 4, Component 2, Investment 1.3 – Partnerships extended to universities, research centers, companies, and research D.D. MUR n. 341 del 15.03.2022 – Next Generation EU (PE0000013 – "Future Artificial Intelligence Research – FAIR'' - CUP: H97G22000210007).

The research of Alfonso Pio Pretorino is supported by a Ph.D. fellowship under the University of Bari Aldo Moro program within the Ph.D. project "Secure Human-Centred Generative AI", co-supported by EULOGIC NT SPA.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] American Speech-Language-Hearing Association, About the american speech-language-hearing association (asha), https://www.asha.org/about/, 2026. Accessed: 2026-03-05.

[2] F. Liégeois, A. Mayes, A. Morgan, Neural Correlates of Developmental Speech and Language Disorders: Evidence from Neuroimaging, Current Developmental Disorders Reports 1 (2014) 215–227. URL: http://link.springer.com/10.1007/s40474-014-0019-1. doi:10.1007/s40474-014-0019-1.

[3] A. Kaspi, M. S. Hildebrand, V. E. Jackson, R. Braden, O. Van Reyk, T. Howell, S. Debono, M. Lauretta, L. Morison, M. J. Coleman, R. Webster, D. Coman, H. Goel, M. Wallis, G. Dabscheck, L. Downie, E. K. Baker, B. Parry-Fielder, K. Ballard, E. Harrold, S. Ziegenfusz, M. F. Bennett, E. Robertson, L. Wang, A. Boys, S. E. Fisher, D. J. Amor, I. E. Scheffer, M. Bahlo, A. T. Morgan, Genetic aetiologies for childhood speech disorder: novel pathways co-expressed during brain development, Molecular Psychiatry (2022). URL: https://www.nature.com/articles/s41380-022-01764-8. doi:10.1038/s41380-022-01764-8.

[4] I. R. Titze, T. Riede, A. Palaparthi, L. S. Hynan, A. Hamilton, L. Toles, T. Mau, Individualized Patient Vocal Priorities for Tailored Therapy, Journal of Speech, Language, and Hearing Research

[1] https://modelcontextprotocol.io/

61 (2018) 2884–2894. URL: http://pubs.asha.org/doi/10.1044/2018_JSLHR-S-18-0109. doi:10.1044/2018_JSLHR-S-18-0109.
[5] K. Pritchard, V. Stojanovik, J. Titterington, E. Pagnamenta, How speech and language therapists and parents work together in the therapeutic process for children with speech sound disorder: A scoping review, International Journal of Language & Communication Disorders 60 (2025) e13132. URL: https://onlinelibrary.wiley.com/doi/10.1111/1460-6984.13132. doi:10.1111/1460-6984.13132.
[6] S. J. Russell, P. Norvig, Artificial intelligence: a modern approach, Prentice Hall series in artificial intelligence, third edition, global edition ed., Pearson, Boston Columbus Indianapolis, 2016.
[7] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language Models are Few-Shot Learners, in: H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems, volume 33, Curran Associates, Inc., 2020, pp. 1877–1901. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.
[8] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer, Journal of Machine Learning Research 21 (2020) 1–67. URL: http://jmlr.org/papers/v21/20-074.html.
[9] M. Calvano, A. Curci, R. Lanzilotti, A. Piccinno, A. Ragone, Leveraging large language models for usability testing: a preliminary study, in: Companion Proceedings of the 30th International Conference on Intelligent User Interfaces, IUI '25 Companion, Association for Computing Machinery, New York, NY, USA, 2025, p. 78–81. URL: https://doi.org/10.1145/3708557.3716341. doi:10.1145/3708557.3716341.
[10] G. Eysenbach, The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers, JMIR Medical Education 9 (2023) e46885. URL: https://mededu.jmir.org/2023/1/e46885. doi:10.2196/46885.
[11] K. Singhal, S. Azizi, T. Tu, S. S. Mahdavi, J. Wei, H. W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl, P. Payne, M. Seneviratne, P. Gamble, C. Kelly, A. Babiker, N. Schärli, A. Chowdhery, P. Mansfield, D. Demner-Fushman, B. Agüera Y Arcas, D. Webster, G. S. Corrado, Y. Matias, K. Chou, J. Gottweis, N. Tomasev, Y. Liu, A. Rajkomar, J. Barral, C. Semturs, A. Karthikesalingam, V. Natarajan, Large language models encode clinical knowledge, Nature 620 (2023) 172–180. URL: https://www.nature.com/articles/s41586-023-06291-2. doi:10.1038/s41586-023-06291-2.
[12] H. Jenkins, Confronting the Challenges of Participatory Culture: Media Education for the 21st Century, The MIT Press, Cambridge, 2009.
[13] B. R. Barricelli, G. Fischer, D. Fogli, A. Mørch, A. Piccinno, S. Valtolina, Cultures of Participation in the Digital Age: From "Have to" to "Want to" Participate, in: Proceedings of the 9th Nordic Conference on Human-Computer Interaction, ACM, Gothenburg Sweden, 2016, pp. 1–3. URL: https://dl.acm.org/doi/10.1145/2971485.2987668. doi:10.1145/2971485.2987668.
[14] M. Calvano, A. Curci, G. Desolda, A. Esposito, R. Lanzilotti, A. Piccinno, Building Symbiotic Artificial Intelligence: Reviewing the AI Act for a Human-Centred, Principle-Based Framework, Minds and Machines 36 (2025) 1. URL: https://link.springer.com/10.1007/s11023-025-09753-w. doi:10.1007/s11023-025-09753-w.
[15] H. Lieberman, F. Paternò, M. Klann, V. Wulf, End-User Development: An Emerging Paradigm, in: H. Lieberman, F. Paternò, V. Wulf (Eds.), End User Development, volume 9, Springer Netherlands, Dordrecht, 2006, pp. 1–8. URL: http://link.springer.com/10.1007/1-4020-5386-X_1. doi:10.1007/1-4020-5386-X_1, series Title: Human-Computer Interaction Series.
[16] G. Fischer, D. Fogli, A. Piccinno, Revisiting and Broadening the Meta-Design Framework for End-User Development, in: F. Paternò, V. Wulf (Eds.), New Perspectives in End-User Development, Springer International Publishing, Cham, 2017, pp. 61–97. URL: http://link.springer.com/10.1007/978-3-319-60291-2_4. doi:10.1007/978-3-319-60291-2_4.
[17] G. Desolda, A. Esposito, F. Greco, C. Tucci, P. Buono, A. Piccinno, Athena: A customizable llm-based code completion tool for visual studio code, in: Companion Proceedings of the 30th

International Conference on Intelligent User Interfaces, IUI ’25 Companion, Association for Computing Machinery, New York, NY, USA, 2025, p. 120–123. URL: https://doi.org/10.1145/3708557.3716357. doi:10.1145/3708557.3716357.
[18] S. Chen, M. Guevara, S. Moningi, F. Hoebers, H. Elhalawani, B. H. Kann, F. E. Chipidza, J. Leeman, H. J. W. L. Aerts, T. Miller, G. K. Savova, J. Gallifant, L. A. Celi, R. H. Mak, M. Lustberg, M. Afshar, D. S. Bitterman, The effect of using a large language model to respond to patient messages, The Lancet Digital Health 6 (2024) e379–e381. URL: http://dx.doi.org/10.1016/S2589-7500(24)00060-8. doi:10.1016/s2589-7500(24)00060-8.
[19] P. Wan, Z. Huang, W. Tang, Y. Nie, D. Pei, S. Deng, J. Chen, Y. Zhou, H. Duan, Q. Chen, E. Long, Outpatient reception via collaboration between nurses and a large language model: a randomized controlled trial, Nature Medicine 30 (2024) 2878–2885. URL: http://dx.doi.org/10.1038/s41591-024-03148-7. doi:10.1038/s41591-024-03148-7.
[20] Y. Du, F. Juefei-Xu, Generative ai for therapy? opportunities and barriers for chatgpt in speech-language therapy, 2023. URL: https://openreview.net/pdf?id=cRZSr6Tpr1S.
[21] American Speech-Language-Hearing Association, About the american speech-language-hearing association, 2025. URL: https://www.asha.org/about/.
[22] Y. Cho, M. Kim, S. Kim, O. Kwon, R. D. Kwon, Y. Lee, D. Lim, Evaluating the Efficacy of Interactive Language Therapy Based on LLM for High-Functioning Autistic Adolescent Psychological Counseling, 2023. URL: https://arxiv.org/abs/2311.09243. doi:10.48550/ARXIV.2311.09243, version Number: 1.
[23] S. Berrezueta-Guzman, M. Kandil, S. Wagner, Integrating AI into ADHD Therapy: Insights from ChatGPT-4o and Robotic Assistants, Human-Centric Intelligent Systems 5 (2025) 230–245. URL: https://link.springer.com/10.1007/s44230-025-00099-1. doi:10.1007/s44230-025-00099-1.
[24] V. Barletta, M. Calvano, A. Curci, A. Pagano, A. Piccinno, Evaluation of ”Speech System” and ”Skill”: An Interaction Paradigm for Speech Therapy:, in: Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies, SCITEPRESS - Science and Technology Publications, 2024, pp. 568–576. URL: https://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0012416700003657. doi:10.5220/0012416700003657.
[25] V. Barletta, M. Calvano, A. Curci, A. Piccinno, A Protocol to Assess Usability and Feasibility of e-SpeechT, a Web-based System Supporting Speech Therapies:, in: Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies, SCITEPRESS - Science and Technology Publications, Lisbon, Portugal, 2023, pp. 546–553. doi:10.5220/0011893300003414.
[26] M. Calvano, A. Curci, A. Esposito, R. Lanzillotti, A. Piccinno, A. P. Pretorino, Leveraging Emotion Recognition to Power Adaptability for More Effective Speech Therapies, in: Joint Proceedings of IS-EUD 2025: 10th International Symposium on End-User Development, volume Vol-3978, CEUR Workshop Proceedings, Munich, Germany, 2025. URL: https://ceur-ws.org/Vol-3978/short-s2-05.pdf.
[27] M. T. James Phoenix, Prompt Engineering for Generative AI, O’Reilly Media, Inc, 2024. URL: https://www.oreilly.com/library/view/prompt-engineering-for/9781098153427/.
[28] S. Shin, J. Oh, S. Lee, Can LLMs See What I See? A Study on Five Prompt Engineering Techniques for Evaluating UX on a Shopping Site, in: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, ACM, Yokohama Japan, 2025, pp. 1–7. URL: https://dl.acm.org/doi/10.1145/3706599.3720079. doi:10.1145/3706599.3720079.
[29] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, in: Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS ’20, Curran Associates Inc., Red Hook, NY, USA, 2020.
[30] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, Technical Report 8, 2019.

[31] A. Z. John Berryman, Prompt Engineering for LLMs, O'Reilly Media, Inc, 2025.
[32] M. Polignano, P. Basile, G. Semeraro, Advanced Natural-based interaction for the ITAlian language: LLaMAntino-3-ANITA, 2024. URL: https://arxiv.org/abs/2405.07101. doi:10.48550/ARXIV.2405.07101.