Exploring Large Language Model as a Design Partner through Verbal and Non-verbal Conversation in Architectural Design Process

Lok Hang Cheung[1], Juan Carlos Dall'Asta[1], Giancarlo Di Marco[1]

[1] Xi'an Jiaotong-Liverpool University, Suzhou, China
Lokhang.Cheung19@student.xjtlu.edu.cn; Juancarlos.Dallasta@xjtlu.edu.cn;
Giancarlo.DiMarco@xjtlu.edu.cn

Abstract. This paper proposes a framework for applying Large Language Models (LLM) as a design partner in architectural design processes instead of a passive question-answering machine. The proposed design framework integrates LLM and Conversation Theory (CT) into a standard parametric design tool for architectural designers. The program establishes an ongoing conversation with the designer through verbal and non-verbal feedback by tracking brain activity and modelling commands. The program can collect conversation data for fine-tuning, thus progressively improving conversation effectiveness. The paper contributes to the knowledge area of architectural design by introducing a novel approach to integrating LLM and CT into the design process, simulated as a proof-of-concept pilot study within a commonly used design software.

Keywords: Artificial intelligence, Large language model, Human-computer interaction, Conversation Theory, Architectural design process

1 Background and Research Framework

1.1 Pandemics of Today's AI and Concern in Design Discipline

With the famous quote from Cedric Price (1979), "*Technology is the answer, but what was the question*", it has been emphasised that despite technology development is rapid and unstoppable, careful attention should be place at "how" they are being applied in the real-world context. Especially with the idea of ubiquitous computing, computational technology including artificial intelligence (AI) is everywhere within nowadays built environment (Weiser, 1991).

Consequence of misunderstanding or abusive use of computing could potentially lead to undesired consequences in the design discipline (Glanville, 1992). More worryingly, they have been gradually happening in the modern society, where the phenomenon was named as "Pandemics of today's AI" by Paul Pangaro (2020).

1.2 Conversation as the Focus of Human-Computer Interaction

Although "Pandemics of today's AI" is happening, there are indications from the cybernetics discipline that offer potential perspectives to tame this situation. Cybernetics originated as a study of "control and communication between animals and machines" (Wiener, 1948), and soon gave rise to the field of Human-Computer Interaction (HCI). HCI has been researched and developed in the architecture discipline focusing on two categories: user-built environment and architect-design software.

Although HCI and system theory encompass a vast range of topics, the architecture discipline has recognised the importance of prioritising conversation as the primary form of interaction, highlighting the interactions between humans and computers in the built environment as a design paradigm (Glynn, 2008; Stralen, 2015; Sweeting, 2021). Its inherent learning process sets conversation apart from other forms of interaction, which holds particular significance in HCI-based architecture due to unpredictable variables, including the use of buildings, environment, end-users, etc. (Pask, 1969).

1.3 The potential of Large Language Models (LLMs)

To achieve meaningful HCI-based design, it is necessary to consider both the theory and technology together (Bateson, 2018). In this paper, a three-step structure is proposed:

(a) If Conversation (learning), then AI (machine learning)

Conversation involves a learning process described above, which conceptually matches the definition of machine learning by Arthur Samuel (1959) "gives computers the ability to learn without being explicitly programmed". However, it is important to point out that "learning" in machine learning refers to a training process, while "learning" in conversation refers to making agreements between individuals through iterative dialogues (Dubberly, Pangaro, 2009). Despite there being technical differences, comparable conceptual similarities are found.

(b) If AI, then Natural Language Processing (NLP)

Natural Language Processing (NLP), a discipline to explore how humans and computers can communicate using human language instead of computer codes, is commonly agreed to be the candidate for HCI design (Alkatheiri, 2022;

Heuer and Buschek, 2021; Wang et al., 2021). NLP exploration in architecture can be traced back to the 1960s. URBAN5, an architectural drawing program, was developed (Negroponte, 1967) by the Architecture Machine Group (AGM). The program would provide verbal feedback on screen using natural languages based on the design drawings and pre-entered design requirements (Negroponte, 1967).

(c) If NLP, then Large Language Models (LLMs)

Large Language Models (LLMs), such as ChatGPT and Bard, have been used widely recently. LLMs are pre-trained models which intend to provide general-purpose applications through question-answering with a basic understanding of sentence structure and knowledge (Ng, 2022). The potential of LLM came from the possibility of Transfer Learning. It is a low-cost technique for individuals to train the LLM with a small amount of specific data to become a domain expert (Howard and Ruder, 2018).

2 Simulation Scenario

2.1 Integration of CT and LLM in a Design Scenario

A The experiment is primarily inspired by Musicolour (Pask, 1971) and URBAN5 (Negroponte, 1967). Musicolour, created by Gordon Pask, was pivotal in developing Conversation Theory (CT). This instrument went beyond simply responding to the musical notes being played by projecting corresponding colours of lighting. It featured a memory mechanism that measured the ranges of tones being played. If a similar range of music notes were played repeatedly, Musicolour would exhibit a sense of "boredom," resulting in less responsive lighting to prompt improvisation from the musician.

On the other hand, URBAN5 was an architectural drawing software developed by Architecture Machine Group (AGM). This program utilised Natural Language Processing (NLP) techniques to provide verbal feedback to designers by analysing computer drawings and pre-entered design requirements (Negroponte, 1967). This experiment attempted to integrate ideas from both case studies into an architectural design scenario. There are two main purposes of the simulation. First, it is to understand the impact of verbal/non-verbal conversations in real-world scenarios. Second, we aim to understand the feasibility of applying LLM to standard architecture software.

In addition to the two case studies, two more points require supplementation. Firstly, there has been the application of NLP techniques in the study of architectural design behaviour recently (Gao et al., 2022, Liu et al., 2022), which shows the potential of NLP is not limited to verbal response as illustrated in the URBAN5. Secondly, CT emphasises the importance of categorising individuals as either mechanical-individuals (m-individuals) or psychological-individuals (p-individuals) (Pask et al., 1973). According to Pask (1980), conversation can only occur when P-individuals are involved. Therefore, in this experiment, we endeavoured to integrate the measurement of brain activity, recognising its relevance within the framework of CT.

Within the simulated scenario, a designer will perform design activities by 3D modelling. The design behaviours and brain activities will be collected and sent to the LLM. Consequently, the LLM provides verbal feedback on the screen and non-verbal feedback through the colour display on a secondary screen, creating a comprehensive interactive experience.

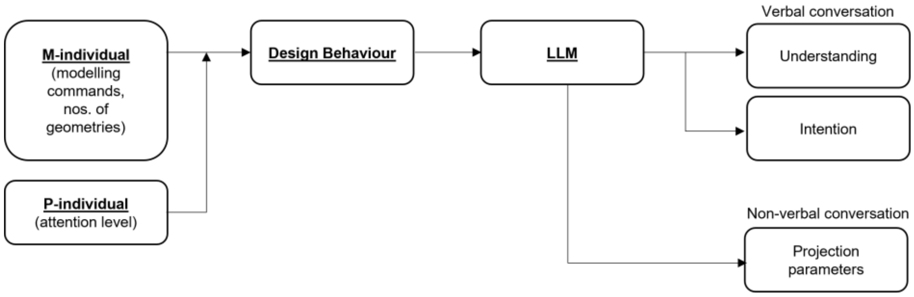

Figure 1. Schematic of the Proposed Prototyping Scenario. Source: Author, 2023

2.2 Methodology: Setup and Rhino-Grasshopper Implementation

In this simulated scenario, the participant uses Rhinoceros, a standard architecture modelling software, for the modelling task. The computational algorithm is also designed within the Grasshopper, integrated into the Rhinoceros environment. The selection of Rhinoceros as the platform for simulation not only reflects the realism of a design task but also benefits from the convenient and immediate connections provided by Grasshopper to modelling commands, external apparatus such as brainwave sensors, and the API for the Large Language Model (LLM).

The workflow is divided into four parts: (1) setting up the environment, (2) collecting data on modelling behaviours and psychological conditions, (3) transferring the data to the LLM, and (4) translating the LLM's responses into verbal and non-verbal feedback.

2.3 Step One: Setup Environment

The setup includes both the physical setup, including the installation of brainwave sensors, and the initial settings in the algorithms to establish connections of the sensors and the LLM with the modelling environment. Other settings, such as time intervals between each feedback, can also be adjusted before the simulation begins.

2.4 Step Two: Collect Data

Regarding the design behaviours, information on m-individual and p-individual are collected. Former includes modelling commands and changes in the number of geometries within a specific time interval, such as 60 seconds. On the other hand, a brainwave-sensing headset is employed. To streamline the experiment, the focus is placed solely on measuring the participants' attention levels. In this simulation, the MindWave Mobile 2 headset developed by NeuroSky is selected as it is commonly chosen for its cost-effectiveness and suitability for stable indoor environments with minimal participant movement (Morshad et al., 2020).

2.5 Step Three: Compile Data to LLM

The collected data is compiled into a pre-designed paragraph for the LLM. This paragraph not only includes the collected data but also provides an example of the expected response format, employing the technique of in-context learning (Dai et al., 2022; Min et al., 2022). This technique guides the LLM's understanding of the textual context and the expected type of conversation. For this experiment, the GPT3.5-davinci model developed by OpenAI is utilised, considering factors such as ease of implementation, cost, and response time.

2.6 Step Four: Translate Responses into Feedback

The responses from the LLM are translated into feedback, which is categorised into verbal and non-verbal forms. Verbal feedback follows a two-step process: first, the LLM describes its understanding and reasoning regarding the current design situation (e.g., the designer is struggling due to a lack of attention), and then it expresses its intended response (e.g., providing a warmer indoor environment to relax the designer and enhance creativity). This approach not only provides deeper insights for designers and researchers to comprehend the logic behind the LLM's responses and decisions but also utilises a technique called Chain of Thoughts, enabling the LLM to deliver step-by-step responses with increased accuracy and logical coherence (Wei et al., 2023). Verbal responses are displayed on the Rhinoceros modelling interface, while non-verbal responses, such as changes in brightness and colour, are projected on a secondary screen within the visible range of the designer.

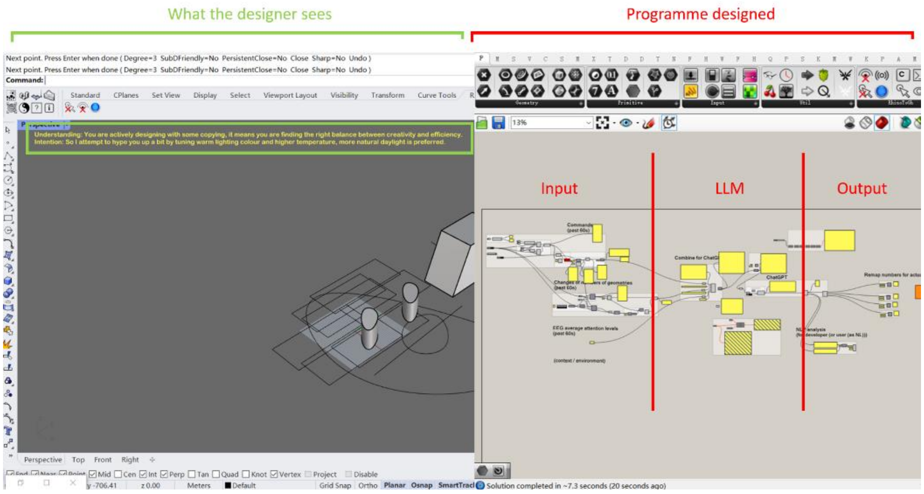

Figure 2. The Modelling Interface in Rhinoceros and Designed Algorithm in the Grasshopper Interface. Source: Author, 2023

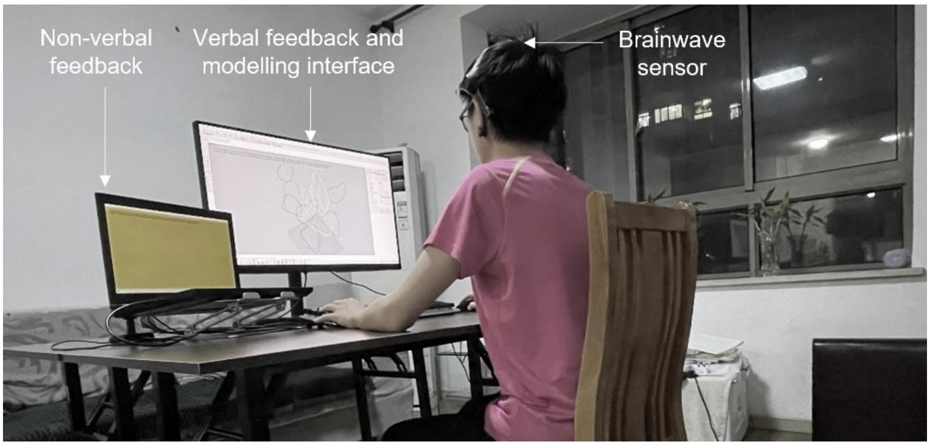

Figure 3. Physical Setup. Source: Author, 2023

3 Findings of Mark I Prototype

3.1 Findings

Using verbal responses in the experiment offers significant advantages by providing direct and explicit sentences to the designer, thereby eliminating the "black box" feeling commonly associated with AI applications. However, certain drawbacks were observed during the study.

Firstly, while the response structure was clear, the content of the responses exhibited inconsistency. Different suggestions were generated when the same input was provided to the LLM for comparison. For example, when the LLM assessed that the designer was making significant design progress, it sometimes recommended calming the designer by displaying dimmer and cooler colours. At other times, it suggested exciting the designer with warmer and brighter colours. This inconsistency raises questions about the reliability and coherence of the LLM's responses.

Secondly, in contrast to the "in the flow" experience reported in the Musicolour case study, the verbal feedback introduced distractions during the design activities. Given that the verbal feedback was updated every minute and required approximately ten seconds to read, it disrupted the flow of the design process and diverted the designer's focus. Similarly, the non-verbal responses presented their own set of distractions. The abrupt changes in colour change every minute interrupted the design activity, resulting in the loss of focus on the modelling tasks. Consequently, the designers had to repeatedly re-focus their attention after each colour change, which displayed counterproductivity compared to the positive impact observed in the Musicolour case study.

3.2 Reflection and Proposed Improvements

The inconsistency observed in the verbal feedback can be attributed to two main factors. Firstly, the design activity is highly subjective, and the LLM was not specifically trained to provide desirable responses tailored to each designer. This lack of personalised training contributes to the inconsistency in the generated responses. Additionally, the inherent probabilistic nature of the LLM further escalated this issue, as it generates different responses even when presented with identical inputs.

To address this problem, a potential solution involves integrating a manual intervention mechanism that allows the designer to express dissatisfaction when undesired responses are received. Collecting the records of adjusted responses, the LLM model can be fine-tuned to become more familiar with the specific designer's behaviours and expected responses. This approach applies the Reinforcement Learning from Human Feedback (RLHF) technique, which combines AI training with human intervention (Ouyang et al., 2022).

Regarding the distraction issue arising from the verbal feedback, one possible solution is to limit its display to situations where designers explicitly seek suggestions or explanations for the LLM's decision on colour changes. Moreover, a "boredom" mechanism, similar to the one employed in the Musicolour case study, can be integrated. This mechanism allows the LLM to evaluate whether the design process is going well, and it would only provide verbal suggestions automatically when the design activity is not progressing smoothly.

In addressing distractions caused by non-verbal responses, it is important to consider selecting response mediums independent of the ongoing design activities. For instance, if the modelling tasks primarily focus on 3D modelling using visual mediums, the response mediums could be related to other spatial qualities such as sound, temperature, or wind. This separation of response mediums from the primary design activities played a significant role in the smooth functioning of the Musicolour case study, where music performances and colour projections utilised different mediums.

Lastly, it is worth noting that after fine-tuning, the LLM model is expected to become more predictable. Consequently, the interaction between the designer and the LLM will shift from conversational to automatic. In this context, the "boredom" evaluation mechanism of the LLM can also adjust the level of response randomness by modifying the "temperature" parameter. OpenAI (2023) defines it as the parameter that controls the randomness and, thus, the creativity of the responses ".

Table 1. Observations from the experiment in the design scenario simulation and proposed improvement taming the problems.

Observed Challenges	**Proposed Improvement**
Distraction from verbal feedback	Only provide necessary verbal feedbacks when the design activity is being evaluated as not going well or upon the designer's request
Inconsistent responses	Allow user intervention for future fine-tuning
Distraction from non-verbal (visual) feedback	Select a non-verbal medium which is independent of the user's activities
Potential problems of predictive, reactive responses after fine-tuning	Add a layer of randomness based on the measurement of creativity of the design activities

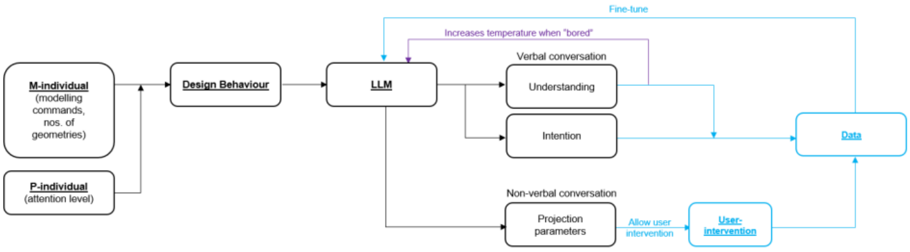

Figure 4. Proposed Schematic for the next improvement. Source: Author, 2023

4 Discussion

4.1 Conclusions and Contributions

The primary objective of this prototyping setup is to demonstrate the practicality of the immediate application of LLM in architecture design scenarios. The proposed design framework, implemented within Grasshopper, presents a novel and practical workflow. It attempted to integrate state-of-the-art LLM technology and CT into a standard parametric design tool, offering architectural designers an innovative approach to their work.

However, there are challenges to overcome on both conceptual and technical levels. Firstly, determining suitable scenarios for conversational built environments is a key consideration. In this case, the experimentation focused on a one-person design desk scenario for rapid prototyping. Yet, it may not directly apply to scenarios that do not require the same level of creativity. As Glynn (2008) illustrated, scenarios like taking a lift will probably not benefit from conversational interactions. Therefore, identifying the appropriate contexts and scenarios for effectively utilising conversational built environments is an important area for further investigation. Also, except for referencing CT-related case studies and integrating the idea of p-individual, the potential of CT has not yet been deeply explored in the experiment.

On the technical side, there are obstacles to address. Each response generated by the LLM upon sending a message took between 6 to 12 seconds, which provided a questionable user experience for the designer. This delay could hinder the fluidity and efficiency of the design process. Resolving these technical challenges is crucial to ensure a realistic and practical application of LLM in real-world architectural scenarios.

4.2 Future Research

As a proof-of-concept pilot study, an architectural design process simulation demonstrates how the entire conversation and fine-tuning workflow could be seamlessly integrated within a standard architecture software environment. This simulation showcases the proposed approach's feasibility and lays the ground for further research and customised applications in architecture design.

In addition to the ongoing Mark 2 prototyping testing with real designers, there are two directions for future development. Firstly, there is a need to explore the transition from a one-to-one scenario to a many-to-many scenario. This expansion would simultaneously facilitate conversations and interactions between multiple designers and the LLM. This exploration would allow for collaborative design processes and the evaluation of the LLM's performance in group settings, thereby broadening the scope of its applicability.

Secondly, there is a need to extend the exploration beyond lighting and examine other spatial qualities such as sound, temperature, and various architectural elements. By incorporating these additional aspects into the conversational framework, designers can engage in meaningful discussions and receive feedback not only on lighting but also on other crucial elements of the built environment. This expansion would enable a more holistic and comprehensive design process, encompassing facades, walls, floors, furniture, and beyond.

Acknowledgements. This research is supported by XJTLU Postgraduate Research Scholarship (PGRSB211206) offered by Xi'an Jiaotong-Liverpool University.

References

Alkatheiri, M. S. (2022). Artificial intelligence assisted improved human-computer interactions for computer systems. *Computers and Electrical Engineering*, 101, 107950. https://doi.org/10.1016/j.compeleceng.2022.107950

Bateson, M.C. (2018) How to be a systems thinker: a conversation with Mary Catherine Bateson. *Edge*. https://www.edge.org/conversation/marycatherine_bates on- how-to-be-a-systems-thinker

Dai, D., Sun, Y., Dong, L., Hao, Y., Sui, Z., & Wei, F. (2022). *Why Can GPT Learn In-Context? Language Models Secretly Perform Gradient Descent as Meta-Optimizers* (arXiv:2212.10559). arXiv. https://doi.org/10.48550/arXiv.2212.10559

Dubberly, H., & Pangaro, P. (2009). What is conversation? How can we design for effective conversation. *Interactions Magazine, 16*(4), 22-28.

Gao, W., Zhang, X., Huang, W., & Shi, S. (2022). Command2Vec: Feature Learning of 3D Modeling Behavior Sequence—A Case Study on "Spiral-stair". In P. F. Yuan, H. Chai, C. Yan, & N. Leach (Eds.), *Proceedings of the 2021 DigitalFUTURES* (pp. 45–54). Springer.

Glanville, R. (1992). CAD Abusing Computing. *CAAD Instruction: The New Teaching of an Architect? [ECAADe Conference Proceedings]* Barcelona (Spain) 12-14 November 1992, Pp. 213-224. http://papers.cumincad.org/cgi-bin/works/paper/cba7

Glynn, R. (2008, March 28). *Conversational Environments Revisited* [Proceedings paper]. In: Kybernetes. Emerald Group Publishing Limited (2008); Emerald Group Publishing Limited. https://doi.org/10.1108/k.2008.06737aab.006

Heuer, H., & Buschek, D. (2021). *Methods for the Design and Evaluation of HCI+NLP Systems* (arXiv:2102.13461). arXiv. https://doi.org/10.48550/arXiv.2102.13461

Howard, J., & Ruder, S. (2018). *Universal Language Model Fine-tuning for Text Classification* (arXiv:1801.06146). arXiv. https://doi.org/10.48550/arXiv.1801.06146

Liu, Q., Gao, W., Huang, Z., & Huang, W. (2022). 3D modeling Logs Based Design Process Mining Method and Its Application for Design Education. *Training, Education, and Learning Sciences*, 59(59). https://doi.org/10.54941/ahfe1002384

Min, S., Lyu, X., Holtzman, A., Artetxe, M., Lewis, M., Hajishirzi, H., & Zettlemoyer, L. (2022). *Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?* (arXiv:2202.12837). arXiv. https://doi.org/10.48550/arXiv.2202.12837

Morshad, S., Mazumder, Md. R., & Ahmed, F. (2020). Analysis of Brain Wave Data Using Neurosky Mindwave Mobile II. *Proceedings of the International Conference on Computing Advancements*, 1–4. https://doi.org/10.1145/3377049.3377053

Negroponte, N. P. (1967). Urban 5—An on-Line Urban Design Partner. *Ekistics, 24*(142), 289–291.

Ng, A (2022) *Deep Learning Specialization* [MOOC]. Coursera. https://www.deeplearning.ai/courses/deep-learning-specialization/

OpenAI (2023). *GPT – OpenAP API*, accessed 11 August 2023. https://platform.openai.com/docs/guides/gpt

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback* (arXiv:2203.02155). arXiv. https://doi.org/10.48550/arXiv.2203.02155

Pangaro, P. (2020) Cybernetics, AI, and Ethical Conversations. AiTech Agora Series, TU Delft, viewed 10 August 2023, https://www.youtube.com/watch?v=VvJpkqKlv9Q&ab_channel=AiTech-TUDelft

Pask, G. (1980). Developments in conversation theory—Part 1. *International Journal of Man-Machine Studies, 13*(4), 357–411. https://doi.org/10.1016/S0020-7373(80)80002-2

Pask, G., Scott, B. C. E., & Kallikourdis, D. (1973). A theory of conversations and individuals (Exemplified by the Learning Process on CASTE). *International Journal of Man-Machine Studies, 5*(4), 443–566. https://doi.org/10.1016/S0020-7373(73)80002-1

Pask, G. (1971) "A Comment, A Case History, and a Plan", *Cybernetics, Art and Ideas*, Reichardt, J., (Ed.) Studio Vista, London, pp.76-99

Pask, G. (1969) The Architectural Relevance of Cybernetics, *Architectural Design*, September issue No 7/6, John Wiley & Sons Ltd (London), pp 494-6

Price, C. (1979). Technology is the answer but what was the question?. [Sound Recording]. Retrieved from PidgeonDigital: https://www.pidgeondigital.com/talks/technology-is-the-answer-but-what-was-the-question-/

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development, 3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Stralen, M. van. (2015). The machine for living in the conversational age. *Kybernetes, 44*(8/9), 1388–1396. https://doi.org/10.1108/K-11-2014-0241

Sweeting, B. (2021). Conversation, fun, and boredom: Cybernetic approaches to intelligent environments in the work of Gordon Pask. Conference: Public Understanding of AI Seminar Series, Centre for Digital Media Cultures, University of Brighton, Brighton. https://research.brighton.ac.uk/en/activities/conversation-fun-and-boredom-cybernetic-approaches-to-intelligent

Wang, Z. J., Choi, D., Xu, S., & Yang, D. (2021). Putting Humans in the Natural Language Processing Loop: A Survey (arXiv:2103.04044). arXiv. https://doi.org/10.48550/arXiv.2103.04044

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models* (arXiv:2201.11903). arXiv. https://doi.org/10.48550/arXiv.2201.11903

Weiser, M. (1991) The Computer for the 21 st Century. *Scientific American, 265*(3), 94–105. Available from: http://www.jstor.org/stable/24938718 (Accessed: 10 August 2023)

Wiener, N. (1948) Cybernetics: Or, Control and Communication in the Animal and the Machine. MIT University Press, Cambridge, MA, 212.