AI and Prompt Architecture – A Literature Review

Cassandra Ansara
Independent Researcher
Not Disclosed
Dawsonville, GA 30534

ABSTRACT
Prompt Architecture represents a novel and systematic approach to the design and optimization of prompts within Conversational AI systems. This literature review synthesizes key developments, methodologies, and insights in the field, drawing from historical influences, recent advances, and current challenges. The review begins with an examination of early influences, such as Weizenbaum's ELIZA chatbot and Minsky's Frames Paradigm, and proceeds to explore modular prompting strategies, optimization techniques, and evaluation methods. Attention is given to innovative approaches, applications in conversational systems, user-centered design, knowledge representation, and ethical considerations. The review identifies existing gaps in the field, including the need for standardized benchmarks, inclusiveness, and ethical oversight. It concludes with a set of recommended actions for further research and development. The insights and recommendations provided in this review contribute to the maturation of Prompt Architecture as a robust and ethical methodology, with potential implications for the broader field of language model interaction and design.

General Terms
Artificial Intelligence, Language Models, Conversational Systems, Prompt Design, Knowledge Representation, Ethical Considerations, Optimization Techniques.

Keywords
Prompt Architecture, Modular Prompting, Chain-of-Thought Prompting, Prompt Optimization, User-Centered Design, Large Language Models (LLMs), ELIZA Chatbot, Minsky's Frames Paradigm, Unified Text-to-Text Approach, Accessibility in AI, Ethical Implications of LLMs, Evaluation of Prompt Quality, Weak Supervision in AI, Automated Prompt Generation, Conversational AI Applications.

1. INTRODUCTION
Prompt architecture represents an important emerging field that aims to systematically design and optimize prompts for improved interactions with large language models (LLMs). As conversational AI systems based on LLMs like ChatGPT have demonstrated remarkable capabilities, a key challenge is crafting high-quality prompts to elicit reliable and relevant responses. Prompt architecture provides a principled framework to address this challenge through modular, optimized, and ethical prompt design. By engineering prompts at a systems level, prompt architecture has the potential to unlock the capabilities of LLMs for a wide range of applications. However, developing a robust methodology requires synthesizing insights across prompt optimization techniques, knowledge representation strategies, evaluation metrics, and ethical considerations. This literature review aims to provide a comprehensive overview of the state of prompt architecture and recommend directions to mature it into an effective and responsible methodology.

The review begins by examining early influences that laid conceptual foundations for prompt architecture, including seminal work by Weizenbaum, Winograd, and Minsky. It then explores recent advances in modular prompting strategies and prompt optimization methods. Attention is given to techniques for evaluating prompt quality, applications in conversational systems, and considerations around knowledge representation and ethics. The review synthesizes insights and gaps from across these areas to provide strategic conclusions and recommendations for advancing the field of prompt architecture in a robust and responsible manner.

2. AI & PROMPT ARCHITECTURE
Architecture refers to the design framework and methodology for developing modular, optimized, and ethical systems to enhance interactions in apps Prompt architecture represents an important emerging field that aims to systematically design and optimize prompts for improved interactions with large language models (LLMs). As conversational AI systems based on LLMs like ChatGPT have demonstrated remarkable capabilities, a key challenge is crafting high-quality prompts to elicit reliable and relevant responses. Prompt architecture provides a principled framework to address this challenge through modular, optimized, and ethical prompt design. By engineering prompts at a systems level, prompt architecture has the potential to unlock the capabilities of LLMs for a wide range of applications. However, developing a robust methodology requires synthesizing insights across prompt optimization techniques, knowledge representation strategies, evaluation metrics, and ethical considerations. This literature review aims to provide a comprehensive overview of the state of prompt architecture and recommend directions to mature it into an effective and responsible methodology.

3. FOUNDATIONS & INFLUENCES
3.1 Early Influences
The groundwork for Prompt Architecture is laid by several pivotal studies within AI Systems. Weizenbaum demonstrated the potential of scripted prompt-response rules in the ELIZA chatbot system, pioneering a method of mapping input to responses [23]. Winograd made a significant connection between prompts and automated reasoning, proposing procedural representations [25]. Similarly, Schank and Abelson innovated script knowledge representations activated by prompts [17]. Grice's work in "Logic and Conversation" and Minsky's Frames Paradigm have added further depth to understanding conversational interfaces [3, 12]. Influential early work like Lieberman established conceptual foundations for using prompts to improve end user interfaces and accessibility of AI systems, foreshadowing modern approaches [9].

Table 1. Key Foundational Concepts

Author(s) & Year	Work/Concept	Contribution & Relevance
Weizenbaum [23]	ELIZA Chatbot	Pioneered scripted prompt-response rules; set the stage for modern techniques.
Winograd (1971)	Automated Reasoning	Connected prompts to automated reasoning through procedural representations.
Schank and Abelson [17]	Script Knowledge Representations	Innovated script knowledge activated by prompts; modeled goal-oriented behavior.
Grice [4]	"Logic and Conversation"	Seminal ideas shaping understanding of meaning construction; laid principles aligning with Prompt Architecture.
Minsky [12]	Frames Paradigm	Concept sharing profound connections with Prompt Architecture's goals, though not directly focused on prompts.
Lieberman [9]	End User Interfaces & Accessibility	Established conceptual foundations for using prompts to improve AI system accessibility; foreshadowed modern approaches.
Rahma Chaabouni et al. [16]	Emergent Communication between Neural Agents	Examined compositional structure and generalization in language models; highlighted systemic compositionality's role in improvement.

4. MODULAR PROMPTING

Modular prompting involves breaking down complex reasoning tasks into manageable sub-tasks. This approach has shown substantial improvements in various areas and can be seen in methods such as Decomposed Prompting by Khot et al. and chain-of-thought prompting by Wang et al. [5, 21].

4.1 Recent Advances

A technique has been provided by Chen et al., which leverages weak supervision and pre-trained embeddings [2]. Despite promising results, there might be limitations regarding comprehensive evaluation or potential constraints in using weak supervision for specific tasks. The unified text-to-text work by Raffel et al. also plays a part in modular design, inspiring innovations in consistent and reusable prompting strategies [15].

Recent advances like the zero-shot approach by Kojima et al. & chain of thought prompting approach by Wei demonstrate that large language models have inherent zero-shot reasoning abilities that can be harnessed through careful prompt design [6, 22]. The innovations by Kojima et al. provide evidence that LLMs can perform human-like few-shot inference when constrained to the appropriate knowledge context using prompting. Similarly, the chain-of-thought prompting approach by Wei et al. shows LLMs' latent capacities for multi-step reasoning when guided by modular prompt sequences.

In addition, Yao et al. have introduced the Tree of Thoughts framework that generalizes chain-of-thought prompting for more complex deliberative reasoning and exploration over multiple paths [27]. The innovations by Kojima et al. provide evidence that LLMs can perform human-like few-shot inference when constrained to the appropriate knowledge context using prompting. Similarly, the chain-of-thought prompting approach by Wei et al. shows LLMs' latent capacities for multi-step reasoning when guided by modular prompt sequences.

Overall, techniques like zero-shot prompting and chain-of-thought prompting demonstrate that modular prompts can unlock systematic reasoning in LLMs, aligning with cognitive science concepts like Minsky's frames [12]. But designing the reasoning chains still requires human insight and expert knowledge. While automating this process is difficult, modular prompting shows promise in areas like mathematical word problems, logical puzzles, and multi-step procedures when constraints are provided through prompt sequences.

4.2 Strategic Conclusions

Recent advances demonstrate the significant potential of modular prompting techniques to unlock systematic reasoning abilities in large language models. Approaches like zero-shot prompting and chain-of-thought prompting provide evidence that LLMs have strong latent capacities for multi-step inference, deduction, and deliberative problem solving. With further refinements in robustness and transferability, modular prompting is poised to enable models to dynamically combine reasoning chains to address novel tasks and scenarios.

However, challenges remain in thorough engineering prompts to be resilient against fragility from small wording variations. Developing standardized prompt modules and interfaces could improve robustness and transfer learning. There are also open questions around optimally balancing human oversight and automation when constructing the modular prompt sequences. Achieving the right blend of human insight and machine efficiency will be key.

Overall, the field of modular prompting is rapidly evolving with innovations in eliciting complex reasoning from LLMs. This progress will lead to more flexible systems that can update beliefs and goals as new information is introduced. Such adaptive and explainable reasoning aligns with the objectives of robust and transparent Prompt Architecture. But realizing the full potential will require interdisciplinary collaboration between experts in AI, cognitive science, human-computer interaction, and ethics. With concerted efforts, modular prompting can mature into a framework for safely unlocking the reasoning capacities of large language models while prioritizing human values.

5. OPTIMIZING PROMPTS

5.1 Studies on Prompt Optimization

Optimizing prompts has gained attention in recent research. Researchers from OpenAI have introduced an influential two-step training procedure combining unsupervised pre-training and supervised fine-tuning [14]. This provides important insights into optimizing prompts by leveraging inherent knowledge gained during pre-training. Shin et al. and Sordoni et al. have focused on automating prompt generation and optimizing prompts in deep learning networks [18, 20]. The exploration of transfer learning by Raffel et al. also sheds light on efficient prompt designs [15]. The AutoPrompt technique introduced by Shin et al. demonstrates automated prompt generation through gradient-guided search [18].

Yang et al. bring hope with the introduction of a trainable prompt encoder to generate dynamic soft prompts. The framework alternates between optimizing the prompt encoder and the underlying LLM. It does not require manually designing each prompt, enabling more automated prompt engineering. However, clear limitations around reliance on a fixed prompt template and lack of analysis into what makes an effective prompt. The dynamic prompting framework contributes towards more automated and efficient prompt tuning. It also demonstrates prompts can be optimized as trainable components integrated with the LLM.

Yao et al. have proposed the ReAct paradigm that interleaves acting and reasoning for task completion [28]. This demonstrates the potential of prompting language models to generate contextual actions and reasoning chains. Musker and Pavlick have provided valuable analysis on how GPT models understand word meaning, shedding light on how to design prompts that elicit accurate reasoning [13].

Furthermore, Zhou et al. provide evidence that large language models can perform at a human level when used for prompt engineering [29]. They underscore the significant impact of careful prompt design.

5.2 Innovative Approaches

Several innovative prompting techniques have been proposed but require further analysis and experimentation. Liu et al. (2021) have introduced P-tuning, optimizing continuous prompts for GPTs to enhance NLU. However, an in-depth analysis of the factors behind effective prompts is lacking. Prompt tuning by Lester et al. leverages soft, learnable prompts, but does not explore limitations or challenges [7].

6. PROMPT QUALITY

6.1 Current Methods & Challenges

Evaluating prompt quality is still an underdeveloped area. Li et al. have made strides with peer-based discussion and ranking but face scalability and disagreements [8]. Standardized, universal metrics for prompt evaluation are in demand.

Zhou et al. (2022) have provided valuable insights by demonstrating the capabilities of LLMs as human-level prompt engineers. Their work emphasizes the significant impact of prompt design on model performance. However, further research is needed on adapting such approaches across diverse tasks.

6.2 Evaluating Prompt Quality

Evaluating prompt quality remains an underdeveloped area and a key challenge. There is a need for standardized, universal metrics to systematically assess and compare the quality of different prompts and prompting approaches. Peer-based discussion and ranking methods have been explored, but face issues of scalability and disagreements. Developing quantitative quality metrics tailored for prompts is an open problem. Factors like relevance, coherence, accuracy, and completeness could be considered. But associating quality scores with diverse conversational prompts poses difficulties. Testing prompts in realistic settings and documenting their limitations provides additional insights beyond constrained evaluations. Overall, systematically benchmarking prompt quality to guide optimization is an active area of research within the broader field of prompt engineering.0

7. REAL WORLD APPLICATIONS

Researchers have started exploring applications of prompt architecture in conversational systems like chatbots. However, most testing has occurred in constrained environments. Thorough real-world evaluations are needed to understand trade-offs and challenges compared to controlled settings.

7.1 Applications & Findings

The Persona prompt pattern adopts perspectives to shape LLM responses but risks incorrect assumptions [24]. The paper does not explore more implicit techniques for controlling dialog beyond explicit prompts. An analysis of different prompting strategies could further inform prompt design for controllability. The benchmark is also limited to written dialog, and does not address challenges in spoken dialog systems. The CommonGen benchmark proposed by Lin et al. represents a valuable testbed for evaluating dialog generation controllability using prompting [10]. An analysis of different prompting strategies could further inform prompt design for controllability. The benchmark could be used to develop and evaluate prompt design patterns aimed at controlling dialog generation in conversational AI systems.

The P-tuning method proposed by Liu et al. for GPTs has implications for conversational systems [11]. By optimizing continuous prompts, P-tuning not only improves NLU performance but also demonstrates the latent knowledge in pre-trained models, which could enhance conversational AI capabilities. The findings of Liu et al. underscore the importance of innovative prompt engineering for unlocking the full potential of language models [11]. However, The paper does not provide an in-depth analysis of the factors that make certain continuous prompts more effective than others, and only limited ablative experiments are conducted to understand the impact of different P-tuning design choices. Comparisons on a broader range of NLU datasets and real-world applications could further demonstrate effectiveness.

Real-world testing can provide further insights into the limitations and trade-offs of using prompt programmed LLMs in conversational contexts.

7.2 Risks of Incorrect Assumptions with Personas

The Persona prompt pattern adopts perspectives to shape LLM responses but risks incorrect assumptions [24]. In complex conversations, personas may lead to inappropriate or inconsistent responses if applied rigidly without adaptations. For instance, conversations often rapidly shift in unpredictable ways. Sticking to a predefined persona could generate responses that do not fit the actual context.

7.3 Limitations of Smart Reply

The introduction of the Smart Reply system by Kannan et al. demonstrates a practical application of prompt architecture principles to email communication, providing contextually relevant automated response suggestions [4]. However, limitations around relevance and scalability would arise if deploying Smart Reply more broadly beyond emails. The word cues effective for emails would often lead astray in noisy social media conversations. Generating high-quality, nuanced responses at scale remains difficult.

7.4 Bridging Research and Applications

More open collaborations between researchers and industry practitioners are needed to close the gap between academic prompt innovations and viability in real conversational systems and products. Insights from large-scale deployments would accelerate translating promising ideas into robust solutions. Ongoing knowledge sharing and testing across domains can help evolve prompt architecture into an impactful methodology for complex conversations.

8. USER CENTERED PROMPT DESIGN

Lieberman's early research introduced core principles of using prompts for intuitive interfaces, aligning with goals of human-centered and accessible prompt design [9]. He introduced the concept of prompts as a way for end users to communicate goals and domain knowledge to AI applications. This pioneering work introduced prompts as a technique to improve usability of AI systems, establishing critical foundations for the field of prompt architecture. Lieberman argues that the user interface is critical for unlocking the utility of AI technologies. Prompts serve as an intuitive interface enabling end users to leverage AI capabilities for their own goals, without needing programming expertise. This helps fulfill key criteria for learnability, predictability, and responsiveness in interfaces. These key principles of simplifying human-AI communication and making AI accessible to end users with intuitive interfaces remain highly relevant to modern prompt engineering.

9. KNOWLEDGE REPRESENTATION

9.1 Frames

Minsky introduced the foundational concept of frames as a tool for knowledge representation [12]. Frames serve as structured representations of stereotypical situations, with "terminals" that can be filled with specific details. This aligns closely with the goals of modular and composable prompting. Prompt modules can be seen as attaching to the terminals of frames, providing the constraints to activate different knowledge contexts. For example, a "restaurant" frame could have slots for prompts specifying cuisine type or pricing. Modular prompts filling these slots would systematically guide an LLM's reasoning within the given restaurant domain. Advanced prompt architectures could potentially incorporate larger frame-based knowledge graphs.

9.2 P Tuning

Liu et al.'s P-tuning technique represents another leap in knowledge representation [11]. P-tuning optimizes continuous prompt embeddings to unlock additional knowledge captured during an LLM's pre-training phase. This reveals LLMs encode more understanding than is directly accessible through standard prompting. Well-designed continuous prompt embeddings activate relevant latent knowledge to handle novel tasks. The ability of techniques like P-tuning to update and leverage knowledge encoded in an LLM's parameters underscores the importance of innovative prompt engineering. It demonstrates prompts can reveal more of the knowledge structures inherent in language models.

9.3 Other Notable Contributions

Si et al. have also advanced knowledge representation by developing prompts to improve GPT-3's reliability across facets like generalization and factuality [19]. Their techniques demonstrate prompting's ability to update LLMs' capabilities by optimizing continuous prompts. This shows language models capture more knowledge during pre-training than previously assumed, highlighting the pivotal role of prompt engineering in leveraging inherent knowledge structures. The paper's focus on reliability also aligns with the goal of Prompt Architecture to create a user experience that is not only engaging but also reliable and trustworthy. They also release all processed datasets, evaluation scripts, and model predictions, which can be valuable resources for other researchers and practitioners in the field.

10. ETHICAL IMPLICATIONS

The ability of prompts to shape LLM content generation raises ethical risks if not addressed responsibly. Biases could be inadvertently encoded into prompts, leading to discriminatory and harmful model behavior. The lack of transparency around proprietary prompts exacerbates such concerns. To mitigate hazards, prompt programming should be treated as a design choice with ethical implications for model outputs. Prompt frameworks must enable transparent auditing and alignment with social values. Researchers have called for increased oversight of LLMs given their societal influence. Developing prompt architecture as an accountable methodology requires incorporating ethical foresight practices. This entails continuous risk assessment, documentation, and monitoring of potential harms. Concrete steps like ethical data sourcing, bias testing prompts, and enabling public scrutiny can help make prompt-based systems trustworthy.

LLM transparency and oversight are necessary to address ethical risks like misuse. Prompt programming affects what content LLMs generate. This aligns with concerns raised by Bender et al. on the implications of LLM design choices & whether models can be too big [1]. Prompt architecture research should focus on developing transparent and value-aligned prompting frameworks.

11. CONCLUSION

In summary, this literature review synthesizes key developments in the emerging field of prompt architecture. Early influences like ELIZA and frames provide conceptual foundations, while recent advances demonstrate the potential of methods like modular prompting and optimization techniques. However, gaps remain in areas like standardized evaluation, inclusivity, and ethical oversight. Developing prompt architecture as a robust and responsible methodology requires addressing these challenges. The recommended next steps include creating benchmark tests, incorporating ethical foresight practices, and collaborating across research and application domains. Additionally, further work is needed in leveraging knowledge representation techniques and testing real-world viability. With concerted efforts, prompt architecture can mature into an accountable framework for safely unlocking the capabilities of LLMs. The insights from this review contribute to clarifying the current state of prompt architecture and guiding targeted advancements in this influential area of AI system design.

11.1 Next Actions for Prompt Architecture

This attempt at a literature review has tried to synthesize key insights and open challenges for developing effective and ethical Prompt Architecture systems. Based on the gaps identified across prior studies, here are some recommended next actions to advance this emerging field:

- Research techniques like chain-of-thought prompting that can elicit complex reasoning from language models using carefully designed prompts. However, thoroughly analyze sensitivity to prompt details and engineer robustness.

- Develop standardized benchmarks and quantitative metrics tailored to systematically compare different prompting techniques and architectures. These can facilitate objective evaluations.

- Explore automated prompt optimization methods while also conducting human evaluations to determine the core factors behind high-quality prompts. Balance automation with human-centered design.

- Investigate the principles of the unified text-to-text approach by Raffel et al., such as simplifying language tasks into standard formats and leveraging transfer learning [15]. Exploring these methodologies could inspire innovative ways to structure, optimize, and utilize prompts in Prompt Architecture, contributing to the field's ongoing development.

- Research compositional prompting frameworks that can dynamically chain prompt components to handle complex interdependent tasks. Build upon modular prompting techniques.

- Extend accessible and inclusive prompting to broader contexts beyond current disability-focused applications through participatory design processes.

- Establish frameworks to monitor potential harms of prompt-based systems and enable transparent auditing. Incorporate ethics into the prompt design lifecycle.

- Test prompt architectures in realistic conversational agents and interfaces to better understand trade-offs and challenges compared to constrained settings.

- Share reusable prompt design patterns and document successful techniques through public knowledge bases. Facilitate transferability of insights across domains.

- Develop flexible prompting tools and interfaces enabling easy exploration of prompt variations for continuous refinement. Support prototyping.

Through concerted efforts across these areas, researchers can make significant headway in maturing Prompt Architecture into a robust, ethical, and human-centric methodology for unlocking the capabilities of language models. But a collaborative spirit and commitment to transparency will be key to realizing the full potential.

12. REFERENCES

[1] Bender, E., Mcmillan-Major, A., Shmitchell, S., Gebru, T., & Shmitchell, S.-G. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Timnit Gebru * timnit@blackinai.org Black in AI Palo Alto, CA, USA CCS CONCEPTS • Computing Methodologies → Natural Language processing. ACM Reference Format. https://doi.org/10.1145/3442188.3445922

[2] Chen, M. F., Fu, D. Y., Sala, F., Wu, S., Mullapudi, R. T., Poms, F., Fatahalian, K., & Ré, C. 2020. Train and You'll Miss It: Interactive Model Iteration with Weak Supervision and Pre-Trained Embeddings. ArXiv.org. https://doi.org/10.48550/arXiv.2006.15168

[3] Grice, H. P. 1975. Logic and conversation. Academic Press.

[4] Kannan, A., Kurach, K., Ravi, S., Kaufmann, T., Tomkins, A., Miklos, B., Corrado, G., Lukacs, L., Ganea, M., Young, P., & Ramavajjala, V. 2016. Smart Reply: Automated Response Suggestion for Email. ArXiv:1606.04870 [Cs]. https://arxiv.org/abs/1606.04870

[5] Khot, T., Trivedi, H., Finlayson, M., Fu, Y., Richardson, K., Clark, P., & Sabharwal, A. 2023. Decomposed Prompting: A Modular Approach for Solving Complex Tasks. ArXiv.org. https://doi.org/10.48550/arXiv.2210.02406

[6] Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. 2022. Large Language Models are Zero-Shot Reasoners. ArXiv:2205.11916 [Cs]. https://arxiv.org/abs/2205.11916

[7] Lester, B., Al-Rfou, R., & Constant, N. 2021. The Power of Scale for Parameter-Efficient Prompt Tuning. ArXiv:2104.08691 [Cs]. https://arxiv.org/abs/2104.08691

[8] Li, R., Patel, T., & Du, X. 2023. PRD: Peer Rank and Discussion Improve Large Language Model based Evaluations. https://arxiv.org/pdf/2307.02762.pdf

[9] Lieberman, H. 2009. User Interface Goals, AI Opportunities. AI Magazine, 30(4), 16. https://doi.org/10.1609/aimag.v30i4.2266

[10] Lin, B. Y., Zhou, W., Shen, M., Zhou, P., Bhagavatula, C., Choi, Y., & Ren, X. 2020. CommonGen: A Constrained Text Generation Challenge for Generative Commonsense Reasoning. ArXiv:1911.03705 [Cs]. https://arxiv.org/abs/1911.03705

[11] Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z., & Tang, J. 2021. GPT Understands, Too. ArXiv:2103.10385 [Cs]. https://arxiv.org/abs/2103.10385

[12] Minsky, M. 1981. A Framework for Representing Knowledge. MIT-AI Laboratory Memo 306, June, 1974. https://courses.media.mit.edu/2004spring/mas966/Minsky%201974%20Framework%20for%20knowledge.pdf

[13] Musker, S., & Pavlick, E. 2023. Testing Causal Models of Word Meaning in GPT-3 and -4. ArXiv.org. https://doi.org/10.48550/arXiv.2305.14630

[14] Openai, A., Openai, K., Openai, T., & Openai, I. 2018. Improving Language Understanding by Generative Pre-Training. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

[15] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research, 21(140), 1–67. https://jmlr.org/papers/v21/20-074.html

[16] Rahma Chaabouni, Kharitonov, E., Bouchacourt, D., Dupoux, E., & Baroni, M. 2020. Compositionality and Generalization In Emergent Languages. https://doi.org/10.18653/v1/2020.acl-main.407

[17] Schank, R. C., & Abelson, R. P. 1977. Scripts, Plans, Goals, and Understanding. Lawrence Erlbaum Associates.

[18] Shin, T., Razeghi, Y., Logan IV, R. L., Wallace, E., & Singh, S. 2020. AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. ArXiv.org. https://doi.org/10.48550/arXiv.2010.15980

[19] Si, C., Gan, Z., Yang, Z., Wang, S., Wang, J., Boyd-Graber, J., & Wang, L. 2022. Prompting GPT-3 To Be Reliable. ArXiv:2210.09150 [Cs]. https://arxiv.org/abs/2210.09150

[20] Sordoni, A., Yuan, X., Côté, M.-A., Pereira, M., Trischler, A., Xiao, Z., Hosseini, A., Niedtner, F., & Roux, N. L. 2023. Deep Language Networks: Joint Prompt Training of

Stacked LLMs using Variational Inference. ArXiv.org. https://doi.org/10.48550/arXiv.2306.12509

[21] Wang, Y., & Zhao, Y. 2023. Metacognitive Prompting Improves Understanding in Large Language Models. ArXiv.org. https://doi.org/10.48550/arXiv.2308.05342

[22] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi Quoc, E., Le, V., & Zhou, D. 2023. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models Chain-of-Thought Prompting.

[23] Weizenbaum, J. 1966. ELIZA---a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36–45. https://doi.org/10.1145/365153.365168

[24] White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. ArXiv:2302.11382 [Cs]. https://arxiv.org/abs/2302.11382

[25] Winograd, T. 1971. Procedures as a Representation for Data in a Computer Program for Understanding Natural Language. Dspace.mit.edu. https://dspace.mit.edu/handle/1721.1/7095

[26] Yang, X., Cheng, W., Zhao, X., Yu, W., Petzold, L., & Chen, H. 2023. Dynamic Prompting: A Unified Framework for Prompt Tuning. ArXiv.org. https://doi.org/10.48550/arXiv.2303.02909

[27] Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., & Narasimhan, K. 2023. Tree of Thoughts: Deliberate Problem Solving with Large Language Models. ArXiv.org. https://doi.org/10.48550/arXiv.2305.10601

[28] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. 2023. ReAct: Synergizing Reasoning and Acting in Language Models. ArXiv.org. https://doi.org/10.48550/arXiv.2210.03629.

[29] Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., & Ba, J. 2022. Large Language Models Are Human-Level Prompt Engineers. ArXiv:2211.01910 [Cs]. https://arxiv.org/abs/2211.01910

13. APPENDIX

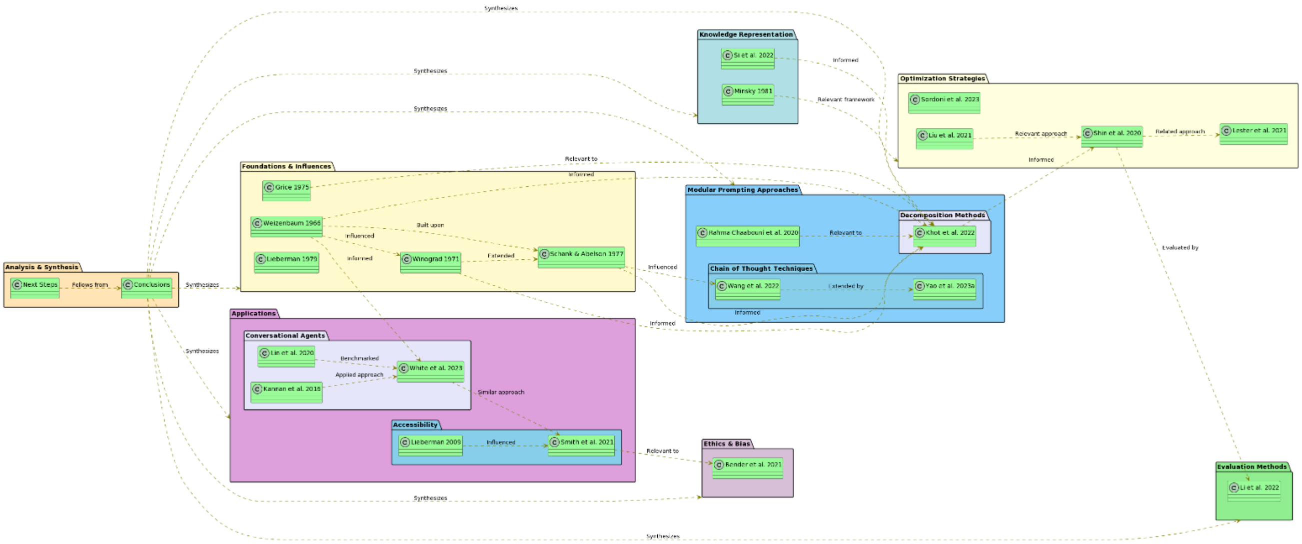

Fig 1: Visual Aid – Prompt Architecture

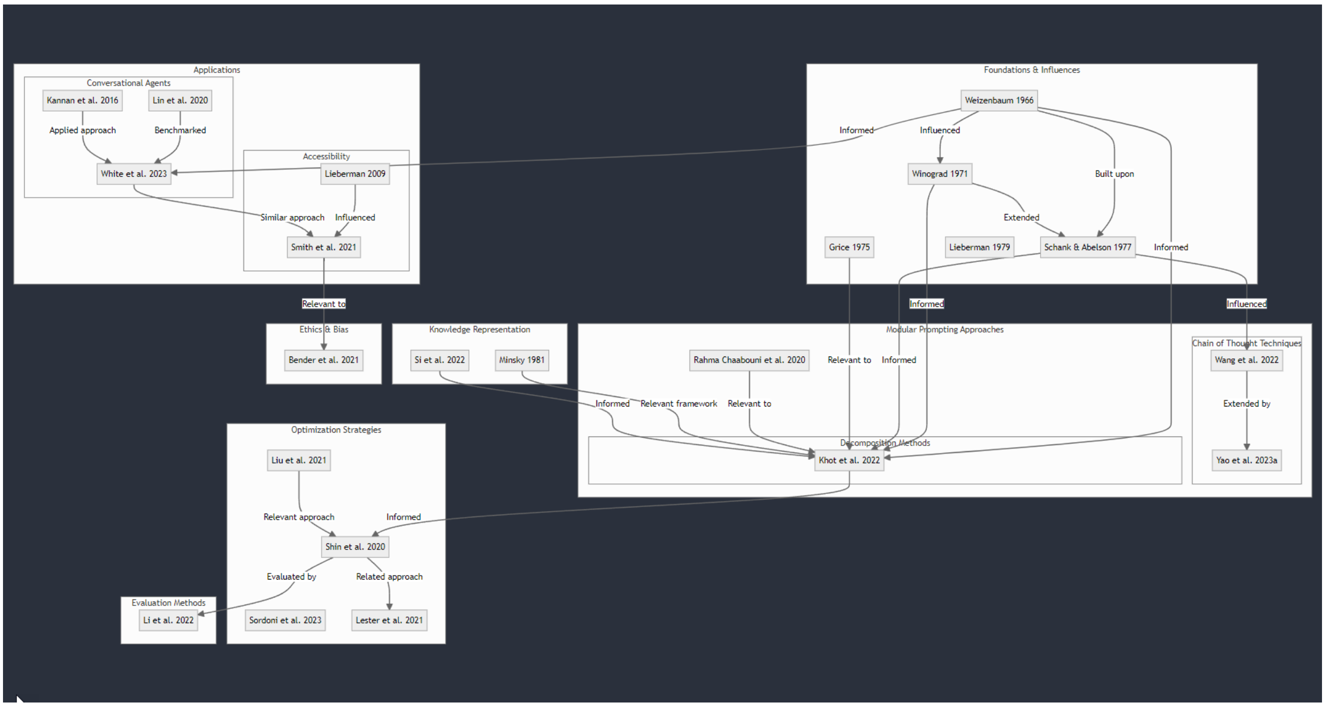

Fig 2: Visual Aid – Literature Analysis