

DOI: https://doi.org/10.48009/1_iis_2026_133

AI prompt engineering enables data visualization: Initial item pool development using a systematic literature review

Assion Lawson-Body, *University of North Dakota*, assion.lawsonbody@und.edu

Jason Jensen, *University of North Dakota*, jjensen@business.und.edu

Laurence Lawson-Body, *University of North Dakota*, laurence.lawsonbody@und.edu

Abdou Illia, *Eastern Illinois University*, aillia@eiu.edu

Kamel Rouibah, *Kuwait University*, kamel.rouibah@ku.edu.kw

Abstract

Generative artificial intelligence (GenAI) and large language models (LLMs) hold significant potential for data visualization across various fields, including business. However, they often generate inaccurate or contextually wrong responses when given vague prompts. This study aims to develop an initial item pool through a systematic literature review to assess the extent to which prompt engineering improves data visualization. Content analysis, combined with the SPAR-4-SLR methodology by Garg & Sharma (2026), was used to develop these measurement tools within structural and measurement frameworks. Of the initial pool, 99 articles that did not meet the inclusion criteria were excluded, along with 26 articles that focused solely on AI or lacked prompt engineering and data visualization. In the end, 86 articles were analyzed in detail. Our findings include: six items assessing the latent variable of data visualization; ten for role prompting; six for direct (zero-shot) prompting; five for one-shot, few-shot, and multi-shot prompting; six for chain-of-thought prompting; six for self-consistency prompting; and seven for knowledge-intensive prompting. Both theoretical and practical insights are provided.

Keywords: prompt engineering, AI, LLMs, data visualization, systematic literature review.

Introduction

Artificial intelligence (AI) has been described as the most quickly adopted business technology of our time (Garry et al., 2024). Major organizations in the financial, pharmaceutical, retail, accounting, marketing, information technology, and entrepreneurship industries have already acknowledged the utility of AI in the workplace and expect their employees, customers, suppliers, and partners to embrace these emerging technologies (Petzel & Sowerby, 2025).

Among the various AI developments and applications, large language models (LLMs) have emerged as a branch of natural language processing (NLP) (Li et al., 2025), leading a transformation with their exceptional language comprehension and generation capabilities (Zhang et al., 2025). GenAI became widely available to the public in 2022 with the release of ChatGPT 3.5 by the San Francisco-based company OpenAI (Akyildiz, 2026). Significant progress in AI has led many major technology companies to develop LLM applications. These companies include OpenAI, Google, and Meta AI, as well as Anthropic and Cohere (Yao et al., 2024). While these companies developed proprietary LLMs (ChatGPT, BERT, BARD,

DialoGPT, BlenderBot), there are open-source LLMs, such as BERT, T5, PaLM, LLaMA, and CTRL (Yao et al., 2024). LLMs excel in diverse linguistic tasks; access to LLMs has proliferated, providing sophisticated tools capable of stimulating detailed and accurate conversations and chats (Petzel & Sowerby, 2025).

LLMs have been applied across various business fields (Xiao & Meng, 2025; Li et al., 2025). LLMs are trained on extensive datasets that can include web-based text, e-books, and other diverse sources. This training process helps LLMs generate AI-created content, such as clear, contextually appropriate text. Responses are produced by predicting the next token in an autoregressive way, guided by learned statistical patterns (Hwang et al., 2025). The quality of these responses depends on factors such as the input prompts. Prompt engineering has become indispensable for LLMs to generate precise, concise, and appropriate responses, all of which are essential elements of data visualization.

In the age of data explosion, creating interactive visualizations is crucial for gaining insights from big data and making informed decisions (Su et al., 2026). However, developing visualizations involves multiple aspects, including data transformation, visual coding, and interaction (Su et al., 2026). To make authoring interactive visualizations easier, various visualization tools such as GGobi, Tableau, visual insight mining (Su et al., 2026), mathematical computation (Chaichana et al., 2024), novel arc-diagram-based visualization (Siddiqui et al., 2026), and Work Trace Visualizer (Mosen et al., 2026) have been developed by both academia and industry (Su et al., 2026). However, there is a scarcity of existing visualization tools that are simultaneously usable, user-friendly, and expressive for visual creators (Su et al., 2026).

Prompts serve as essential tools for interacting with LLMs such as ChatGPT and play a crucial role in determining response quality (Hwang et al., 2025), enabling effective data visualization. However, LLMs sometimes produce inaccurate or nonsensical outputs (known as hallucinations) and, at the moment, often lack common sense and an understanding of reality (Knoth et al., 2024). The lack of clear direction in the trial-and-error process, the poor depiction of activities and their outcomes, concerns about computing costs, and ethical implications are among these problems (Knoth et al., 2024). The 'garbage in, garbage out' slogan has become a reality in the evolution of the NLP field. Well-designed, structured prompt engineering enables LLMs to produce high-quality data visualizations. However, even with advancements, LLMs often produce inaccurate or contextually erroneous responses when prompted with unclear instructions (Hendawi et al., 2026).

Therefore, researchers must answer two questions: Is appropriate prompt engineering the solution to hallucination and bias in LLMs? Does the quality of prompt engineering improve data visualization performance? What are the new instruments for measuring prompt engineering for predicting effective data visualization?

This study aims to develop an initial item pool through a systematic literature review to assess prompt engineering for enabling appropriate data visualization, while considering the hallucination and bias challenges associated with LLMs. More specifically, this study extends the Human-Computer Interaction (HCI) theory to examine the impact of prompt engineering on data visualization. It also aims to conduct a systematic literature review to develop an instrument to measure the success of prompt engineering in data visualization.

Theoretical Background

This research applies HCI theory to explain the relationships between prompt engineering and data visualization. HCI theory draws on many of its characteristics from computer science, information systems,

human resource management, psychology, and ergonomics (Knoth et al., 2024; Garg et al., 2025; Robertson et al., 2024). HCI theory focuses on designing, evaluating, and deploying user-friendly technology systems by examining how humans and machines interact (Garg et al., 2025). Feedback and interactivity are crucial elements in HCI theory because feedback allows real-time responses to user actions. Therefore, HCI is based on how the user interacts with the technology through its interface. HCI uses the PACT (People, Activities, Context, Technology) framework to make computer systems intuitive, efficient, and safe (Knoth et al., 2024). Prompt engineering is more directly tied to fields that focus on interactions between humans and AI, such as HCI, Human-AI Interaction, and conversational AI (Knoth et al., 2024).

Recently, many authors have emphasized the relationship between prompt engineering and HCI. Among them, Garg et al. (2025) found that GenAI and structured prompt engineering training have the potential to transform educational practices and recommend integrating prompt engineering into human-AI collaboration. Prompt design has not yet been studied broadly and systematically from an HCI perspective (Knoth et al., 2024). Prompt construction plays a pivotal role in shaping the interaction between users and generative AI models, serving as the blueprint for communication (Knoth et al., 2024). The last authors added that the prompting methodologies and techniques have implications for understanding and improving the quality of interactions with LLMs from an HCI perspective (Knoth et al., 2024).

On their side, Robertson et al. (2024) used a constructivist theory based on the principle that individuals assimilate new knowledge by bridging it to their existing understanding to develop an AI prompting protocol. They found an active engagement process in the human-AI construction of knowledge through GenAI, thereby enhancing user interaction with AI chatbots. They highlighted that the popularity of ChatGPT, in particular, is partly due to its user-friendly interface, which promotes the model, and its ability to answer questions in a human-like conversational way (Robertson et al., 2024).

Furthermore, Knoth et al. (2024) found that higher-quality prompt engineering skills predict the quality of LLM output and that prompt engineering is indeed a required skill for the goal-directed use of GenAI tools. Knoth et al. (2024) reported that non-experts tended to develop prompts that resembled human-to-human instructions. Building on the work of Knoth et al. (2024), Garg et al. (2025), and Robertson et al. (2024), this study advances HCI theory by exploring how humans and LLMs collaborate through prompt engineering. Its aim is to enhance the adaptability and flexibility of data visualization, supporting human critical thinking, decision-making, and problem-solving.

HCI explains that well-crafted prompts enhance visualization results by facilitating human-AI interactions. According to HCI theory, the effectiveness of these interactions relies on how accurately users convey their intentions to the LLM and how well the system interprets them. Since LLMs are interactive, prompts serve as the main means of communication.

Related Work

Data Visualization

Visualization is the representation of information or data through common visual elements, such as charts, graphs, maps, and animations, often enhanced with interactive capabilities to support dynamic exploration (Lian et al., 2025). Visualization simplifies complex data, making it more accessible and engaging while enhancing trend identification and decision-making (Du et al., 2026). While organizations collect unprecedented amounts of data, its value remains locked without effective interpretation tools (Lian et al., 2025). In this situation, visualization emerges as a powerful medium to bridge the gap between raw data and meaningful knowledge, enabling analysts to perceive patterns, identify anomalies, and communicate findings intuitively (Lian et al., 2025).

There are many tools and software that enable data visualization. First, early tools for interactive data visualization, such as GGobi and Tableau, incorporate common interaction technologies. They use parameter actions, enabling users to interact directly with visualization elements to change parameter values. However, this mechanism can add complexity to simple tasks and increase learning costs for developers (Su et al., 2026).

Second, given the richness and interpretative power of visualizations, visual insight mining – the process of extracting meaningful insights from raw data through intuitive visual representations – has become increasingly vital (Lian et al., 2025). The process of visual insight mining has six distinct tasks (i.e., task decomposition, visualization generation, visualization recommendation, chart parsing, chart question answering, and insight generation) (Lian et al., 2025).

Third, the relevance of data, digitalization, and its transmission continues to increase to stay on the competitive edge (Reisig & Lenz, 2026). For this purpose, the data from a milling machine is captured, processed, and displayed in an appropriate visualization tool (Reisig & Lenz, 2026). The result is a dashboard application that displays and visualizes machine data in real time (Reisig & Lenz, 2026).

Fourth, because visualizing economic growth is challenging and powerful in identifying business performance, Chaichana et al. (2024) developed a mathematical computation to visualize virtual economic growth geographically for small and medium-sized enterprises (SMEs).

Fifth, operational activities are assumed to be totally ordered and hence can be visualized as a sequence of activity labels (Siddiqui et al., 2026). Visualizations largely assume sequential event data, making it difficult to analyze repetitions in partially ordered event data, which better captures real-world process behavior. To address this gap, Siddiqui et al. (2026) present a novel arc-diagram-based visualization that highlights recurring activity patterns across individual process-execution variants. This approach allows analysts to intuitively detect repetitions that are otherwise obscured in raw data or traditional variant views (Siddiqui et al., 2026).

Sixth, Mosen et al. (2026) developed the Work Trace Visualizer, a timeline-based visualization and analysis tool that integrates records from a long-term study to support workgroups and collaborative work. Work Trace Visualizer is an innovative visualization and analysis tool that builds on social documents as traces of asynchronous group work and connects them with ethnographic diary records on synchronous work activities (Mosen et al., 2026).

Collectively, these tools and software require an extensive learning curve, involve human effort, and are connected to data analytics. None of these tools incorporates AI. Some of them are expensive to purchase, making them available only to businesses with resources. Using LLMs and prompt engineering to generate data visualizations offers an alternative approach, enabling businesses to stay competitive.

Prompt Engineering

Central to the design of LLMs is a self-supervised learning objective, often predicting the subsequent words in a sequence or filling in masked words (Hwang et al., 2025). In addition, LLMs are trained on vast datasets that include both accurate information and misinformation, which can sometimes lead to incorrect or biased responses (Garg et al., 2025). To minimize these challenges, one approach is prompt engineering, which involves crafting precise, context-specific queries to guide AI responses (Garg et al., 2025). Prompt engineering requires more than just learning to write clear and purposeful prompts—it also involves the ability to interpret and critically evaluate the responses produced (Akyildiz, 2026).

The influence of prompt engineering extends to numerous disciplines, such as business, social science, and art science (Chen et al., 2026). Prompt engineering relies on LLMs to help people engage and work more effectively and efficiently with these models (Hendawi et al., 2026). Prompt engineering involves more than just asking a simple question to an AI tool; it often requires setting constraints or roles, providing examples, expanding or adjusting the prompt, and critically assessing the model's output (Akyildiz, 2026). To achieve desired results with AI tools, prompt engineering involves carefully choosing words, phrases, and the overall structure of the input (Akyildiz, 2026). There is no consensus on the definition of prompt engineering. Table 1 summarizes the main definitions of prompt engineering.

Table 1. Prompt Engineering Definitions

Definitions	Sources
Prompt engineering is the process of writing instructions that enable a natural language generation tool to perform a task effectively and utilize all its features.	Akyildiz (2026)
Prompt engineering is the process of structuring inputs and has emerged as a crucial technique for maximizing the utility and accuracy of LLMs.	Chen et al. (2026)
Prompt engineering is the art and science of designing and refining effective prompts to elicit the desired output from the LLMs.	Phillips-Wren & Hakansson (2025)
Prompt engineering is an expanding field focused on establishing best practices for creating effective inputs for LLMs to generate desired outputs systematically.	Tani et al., (2025)
Prompt engineering is the practice of developing inputs in a way to elicit the best possible responses from LLMs	Hendawi et al., (2026)
Prompts are instructions given to an LLM to enforce rules, automate tasks, and guarantee certain qualities (and quantities) of output.	Ilagan et al., (2024)
Prompt engineering has evolved into the skill of formulating precise, well-structured instructions to elicit desired responses or information from an LLM, thereby optimizing the effectiveness of the interaction.	Knoth et al., (2024)
Prompt engineering involves crafting the right questions to guide AI in generating relevant, meaningful content that humans can refine for practical application.	Robertson et al. (2024)

Prompt engineering is a skill that entails creating and refining specific inputs for LLMs, enabling users to obtain high-quality outputs from models and to overcome difficulties while leveraging the capabilities of generative AI (Knoth et al., 2024). However, no single evaluation tool can guarantee that businesses use high-quality, reliable prompt engineering. To minimize these challenges, evaluation and measurement instruments for prompt engineering are necessary. This study proposes developing a prompt-engineering-enabled data-visualization evaluation instrument through a systematic literature review.

Prompt Engineering Evaluation and Measurement

As with any AI field, measurement and evaluation are essential for prompt engineering. Recently, several authors have examined prompt-engineering measurement tools. As a matter of fact, prompt optimization was performed by generating evaluations of how well the product and service descriptions matched the customer's e-commerce preferences (Koralewski et al., 2025). In another scenario, the Kruskal–Wallis test was used to determine which prompt-engineering approach produced the easiest-to-read and easiest-to-understand consent forms (Koralewski et al., 2025). In another article, a Boschloo test was used to evaluate causal claims generated by an LLM using real-world intervention-prompt engineering data (Koralewski et al., 2025). Garg et al. (2025) found that structured prompt training effectively enhances students' data analysis and programming skills.

On their side, Endler et al. (2026) offered a relevant study on assessing the effectiveness of AI prompt engineering in enhancing safety management systems. The evaluation uses 30 safety-related prompts across six different operational areas, rated on relevance, clarity, and completeness, with scores from 1 to 5. The findings show a trimodal performance pattern: 60% achieved perfect scores, 30% failed entirely due to indexing and contextual issues, and 10% showed moderate performance with partial correctness or some inaccuracies.

Although previous studies have considered some prompt-engineering evaluation tools, they have largely overlooked how prompt engineering with LLMs produces measurable data visualizations. They did not create measurement tools to evaluate prompt engineering that will produce data visualization. Our study fills these gaps by using a systematic literature review to develop a new measurement to explore the relationships between prompt engineering and data visualization.

Prompt engineering is a method for prompting generative AI models, such as LLMs, built using AI techniques (Phillips-Wren & Hakansson, 2025). The models used in GenAI can produce a variety of outputs, including audio, video, images, or text, based on user input, which is referred to as a prompt (Knoth et al., 2024).

These methods range from trivial prompts to knowledge-intensive prompts. There are role prompting and direct prompting, also known as zero-shot prompting (Phillips-Wren & Hakansson, 2025). Prompting techniques, including at least one example that guides the models' output, are example-based models. For example, one-shot, few-shot, and multi-shot prompting provide one, several, or multiple examples (Phillips-Wren & Hakansson, 2025). Prompt engineering has progressed from simple, structured inputs in the 1950s to advanced methodologies, including self-consistency, chain-of-thought reasoning, and generated knowledge (Chen et al., 2026). Each of these methods is described below.

In data visualization involving large language models (LLMs), these prompting methods differ conceptually because they affect the model via various cognitive processes. Role prompting defines the role the model should assume, while zero-shot and few-shot prompting control the level of guidance through examples. Chain-of-thought and self-consistency prompting shape how the model reasons, and knowledge-intensive prompting adds external or domain-specific knowledge.

Role prompting

Among all prompt engineering methods, role prompting represents a promising approach (Hendawi et al., 2026). In this method, the model is instructed to answer in the context of a role, which improves its behavior and better matches user expectations (Hendawi et al., 2026). An example of a role is "You are a data analyst." Role prompting enhances response understanding and clarity by focusing on the model and setting clear objectives (Hendawi et al., 2026).

There are three role prompts: static, expert, and dynamic. First, static role prompting involves assigning a fixed role to a model to generate output with contextual accuracy and task-specific precision (Chen et al., 2026). This technique is commonly used for tasks needing steady output that matches specific roles (Chen et al., 2026). Second, the expert role prompt method enhances role prompting by producing task-specific identities, demonstrating improved contextual precision and informativeness for specific tasks (Chen et al., 2026). Third, dynamic role-play prompting allows a model to adapt its focus and improve its output as user input changes (Chen et al., 2026). For instance, a follow-up question like "could you elaborate on the economic factors?" directs the model to focus on specific parts of the previous answer, refine its response in real time, and tailor it to the user's evolving needs.

Direct prompting, also known as zero-shot prompting

Zero-shot prompting, where models perform tasks based on general instructions without specific training data (Garg et al., 2025). It provides the model with instructions without examples (Phillips-Wren & Hakansson, 2025). LLMs can solve a variety of NLP tasks zero-shot, without relying on any training data for a given task, simply conditioning the model on appropriate prompts (Ilagan et al., 2024). Zero-shot learning requires prompting LLMs without examples, but it can be enhanced by fine-tuning the instructions and by reinforcement learning via human feedback (Knoth et al., 2024). Related terms to zero-shot are in-context learning (ICL) (Ilagan et al., 2024).

One-shot, few-shot, and multi-shot prompting

Research indicates that prompting techniques such as zero-shot and few-shot prompting enable models to perform new tasks with limited training data, leveraging existing knowledge to provide accurate answers (Garg et al., 2025). Few-shot prompt introductions introduce a few examples to improve model understanding, demonstrating that high-quality examples enhance performance on complex tasks (Garg et al., 2025). Few-shot prompt enables learning from limited examples (Ilagan et al., 2024). Few-shot learning, which trains an LLM to learn a new task with few examples, enables spontaneous task delegation and improves model performance (Knoth et al., 2024).

Initially, a one-shot approach was used by combining all necessary information (including the context and goal) into a single prompt for GPT-3.5 to process (Ilagan et al., 2024). This method only identified a small portion of the expected codes and either misdiagnosed or missed the rest. Because initial tests do not achieve satisfactory accuracy, precision, or recall, prompting should be supplemented with few-shot examples or even fine-tuning of the underlying LLMs (Ilagan et al., 2024).

Chain-of-Thought prompting

The chain-of-thought approach involves providing intermediate reasoning steps to guide a model's responses, a process that can be facilitated by simple prompts (Ilagan et al., 2024; Chen et al., 2026). In other terms, Chain-of-thought entails creating intermediate natural-language reasoning steps to guide LLMs through challenging tasks (Knoth et al., 2024). It also provides a clear structure for the model's reasoning, making it easier for users to understand how the model reaches its conclusions (Chen et al., 2026). This prompt engineering is very close to explainable AI, which addresses the black-box nature of previous AI models. To achieve more accurate and comprehensive diagnoses and encoding, the chain-of-thought prompting method was applied to the problem (Ilagan et al., 2024). The prompts were arranged in three general phases: first, listing of key observations; second, diagnosis; and third, classification (Ilagan et al., 2024).

Chain-of-Thought prompting improves reasoning by breaking down tasks into step-by-step processes, enhancing accuracy in complex problem-solving (Garg et al., 2025). Chain-of-thought prompting advances the technique by enabling step-by-step reasoning, significantly improving performance in reasoning tasks. To automate this process, Automatic Chain-of-Thought prompting was proposed to generate diverse reasoning chains, thereby improving accuracy and robustness without manual effort (Garg et al., 2025).

Self-consistency prompting

It is the consistency in the responses provided by the LLMs. The prompt must be formulated to check for consistency from the LLMs. These types of prompts are recommended in fields that require high precision and proof, such as accounting and medicine. Self-consistency is an advanced prompting technique that aims to ensure model responses are consistent with one another (Chen et al., 2026). Self-consistency, generating multiple reasoning chains, and selecting the most consistent final answer to enhance accuracy through diverse approaches (Garg et al., 2025). This method can involve searching and accessing external sources

relevant to the prompt to obtain new information beyond the LLM's original training dataset for consistency checking (Phillips-Wren & Hakansson, 2025).

Knowledge-intensive prompting

Generate knowledge prompting involves using knowledge or information to help the model make more accurate predictions. It encourages the model to produce outputs and predictions based on factual information or structured knowledge (Phillips-Wren & Hakansson, 2025). The generated-knowledge approach in prompt engineering leverages LLMs' ability to generate potentially useful information about a given prompt before producing a final response (Chen et al., 2026). "Generated knowledge prompting" is a method that improves commonsense reasoning in language models without requiring fine-tuning (Tani et al., 2025). It is a prompting technique that involves using formal logic and reasoning frameworks to solve problems (Phillips-Wren & Hakansson, 2025).

Prompt Engineering and Data Visualization Framework

In this research, the prompt engineering and data visualization framework has two parts. The first part covers prompt engineering methods such as role prompting, direct prompting (also known as zero-shot prompting), one-shot, few-shot, and multi-shot prompting, chain-of-thought prompting, self-consistency prompting, and knowledge-intensive prompting. The second part is data visualization. This framework articulates the relationships between prompt engineering and LLMs, on the one hand, and between prompt engineering and LLM-based data visualization, on the other. Furthermore, the framework serves as a starting point for developing measurement items to evaluate the effectiveness of prompt engineering in enhancing data visualization, as identified through a systematic literature review. Figure 1 shows the research framework.

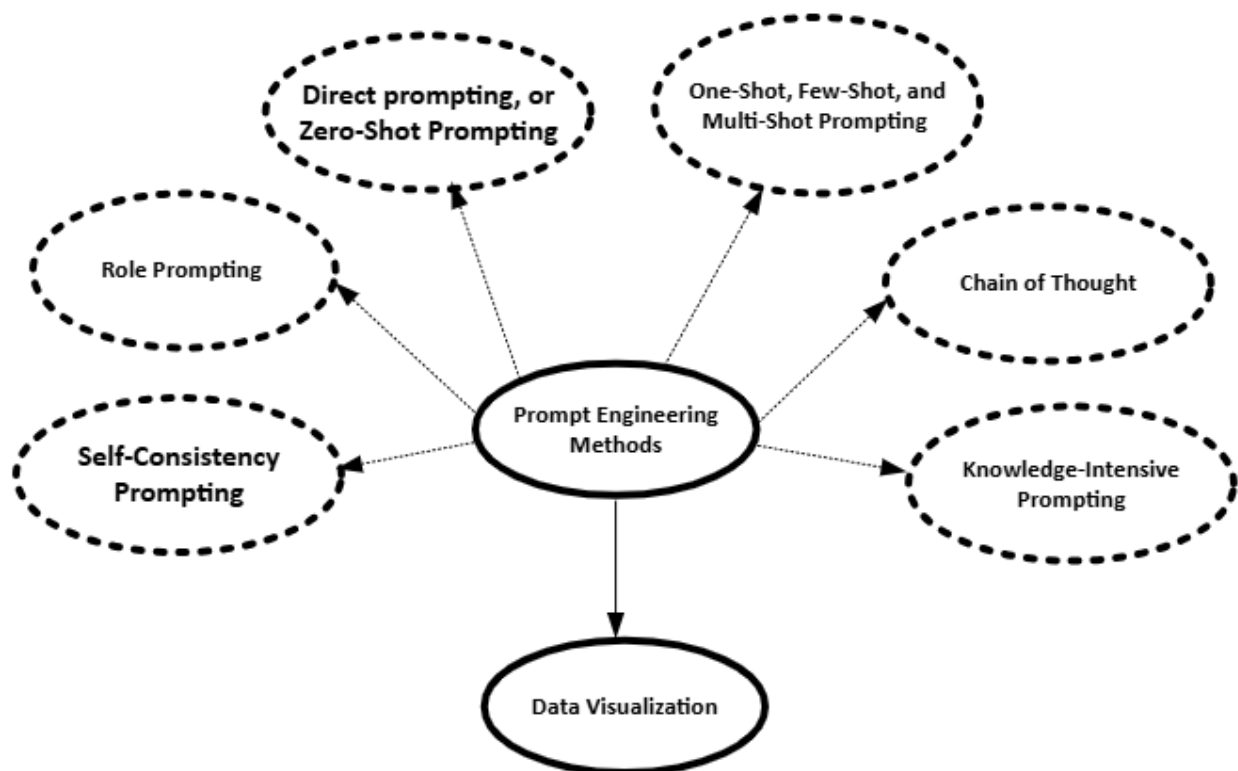


Figure 1. Structural Framework

Research Methodology

Figure 2 summarizes the methodological approach adopted in this study. Our systematic literature review strategy employed qualitative methods, including content analysis, combined with the Scientific Procedures and Rationales for Systematic Literature Reviews (SPAR-4-SLR) methodology by Garg & Sharma (2026). SPAR-4-SLR is a structured framework designed to assist researchers in planning, executing, and reporting systematic literature reviews, thereby improving methodological rigor, transparency, and reproducibility. (Garg & Sharma, 2026). The systematic review of literature is considered one of the most effective means by researchers to delve deeper and obtain a more profound and unbiased grasp of the current literature (Lawson-Body et al., 2024; Garg & Sharma, 2026).

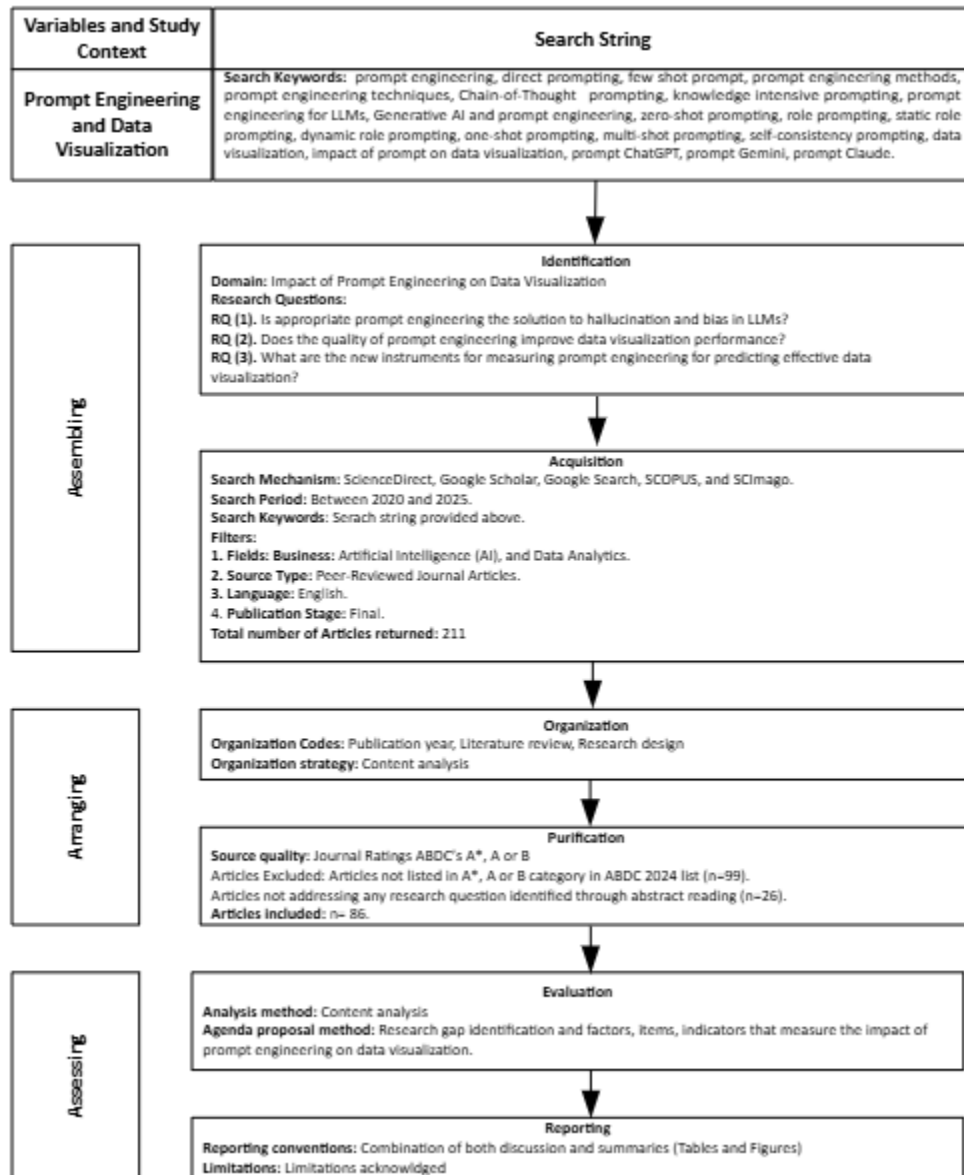


Figure 2. SPAR-4-SLR (Adapted from Garg & Sharma, 2026)

Review Protocol

SPAR-4-SLR by Garg & Sharma (2026) consists of three phases: assembling, arranging, and assessing. Each phase includes two substages. The assembling phase covers identification and acquisition. The

arranging phase focuses on organization and purification based on inclusion and exclusion criteria. The assessing phase involves analysis and reporting.

Assembling

Assembling bears on a) Identification, which outlines the review's scope, domain, research questions, objectives, and key sources; b) Acquisition, which involves finding relevant articles through a well-defined search strategy and keywords.

The research domain is AI and data analytics, as well as the impact of LLM-based prompt engineering on data visualization. The research questions and objectives are announced in the introduction section of this research. As Chonkner et al. (2017) pointed out, the online digital libraries were considered during the acquisition process. The thematic analysis covered peer-reviewed, SCOPUS-indexed, or ABDC (Australian Business Dean Council) journals. More importantly, the thematic keyword search was done in social media using ScienceDirect, Google Scholar, Google Search, SCOPUS, and SCImago. Additionally, The search was guided by the following keywords: prompt engineering, direct prompting, few shot prompt, prompt engineering methods, prompt engineering techniques, Chain-of-Thought prompting, knowledge intensive prompting, prompt engineering for LLMs, GenAI and prompt engineering, zero-shot prompting, role prompting, static role prompting, dynamic role prompting, one-shot prompting, multi-shot prompting, self-consistency prompting, data visualization, impact of prompt on data visualization, prompt ChatGPT, prompt Gemini, prompt Claude. Papers published between 2020 and 2025 were selected. These selections were considered relevant because prompt engineering is new. In total, 211 articles were returned at the end of the “assembling” step.

Studies were included if they (1) appeared in peer-reviewed academic journals; (2) were written in English; (3) directly addressed the research topic in this review; (4) met the established journal quality standards; and (5) provided adequate empirical or conceptual evidence aligned with the review’s objectives. Studies were excluded if they (1) appeared in journals that did not meet the quality thresholds; (2) addressed topics outside this review's scope; (3) consisted of conference papers, book chapters, dissertations, editorials, commentaries, or other non-peer-reviewed outputs; (4) were not published in English; or (5) were duplicates identified during screening. No geographic restrictions were applied, so studies from any country or region were considered.

Arranging

At the organization stage, the initial corpus of the relevant articles on prompt engineering and data visualization was systematically organized using content analysis.

At the purification level, article inclusion was determined using source-quality criteria. A total of 99 articles extracted from journals and reviews that did not meet the ABDC’s A*, A, or B ranking criteria were excluded. Another group of 26 articles that focused solely on AI or did not include prompt engineering or data visualization was removed. Sixteen (16) duplications were found among the 26 articles, which were not addressed by any research questions. Finally, 86 articles were retained for in-depth content analysis and full-text analysis.

The measurement items were developed through a systematic content analysis of 86 studies meeting our inclusion criteria. Each study was meticulously reviewed during coding to identify concepts, features, and practices associated with data visualization and prompt engineering. Similar concepts appearing in multiple studies were grouped into themes, and duplicate or overlapping items were merged. These categories were further refined through iterative comparisons with the literature to guarantee consistency and

comprehensiveness. Therefore, the six data visualization items and the forty prompt engineering items were not predefined but originated from this systematic coding and synthesis process.

Assessing

At this level, the evaluation and reporting of the literature were executed. Content analysis was used to summarize the findings, which include the most influential articles, authors, and journals/reviews. We considered core factors, items, and indicators that measure the impact of prompt engineering on data visualization. Building on these results, potential measurement instruments were created for each prompt engineering methods, and data visualization. Those measurement instruments are reported in Tables 1 through 5 and in Figure 3. Limitations for the research were provided in combination with future research avenues in the conclusion section of this work.

Analysis and Results

Figure 3 shows the measurement framework, including generated items from the systematic literature review, which applied SPAR-4-SLR, adapted from Garg & Sharma (2026). Each item in the measurement should be graded on a scale from 1 (“strongly disagree”) to 7 (“strongly agree”).

The literature review initially identified the conceptual dimensions and their theoretical definitions for each construct. Using these dimensions, we created an initial pool of candidate items by adapting validated scales from earlier studies and tailoring the wording to the current research context. Each item was then evaluated for its representation of the corresponding construct's theoretical and conceptual domain. Redundant, ambiguous, or overlapping items were eliminated, resulting in the final set of measurement items presented in this study. This explanation highlights the logical flow from the literature review through conceptualization to item selection.

We found:

- Six items measuring the latent variable data visualization.
- Ten items measuring the latent variable of role prompting.
- Six items measuring the latent variable direct prompting, also known as zero-shot prompting.
- Five items measuring the latent variable of one-shot, few-shot, and multi-shot prompting.
- Six items measuring the latent variable chain-of-thought prompting.
- Six items measuring the latent variable self-consistency prompting.
- Seven items measuring the latent variable knowledge-intensive prompting.

Tables 1 through 5 also explain how each factor or latent variable is connected to its measurement items.

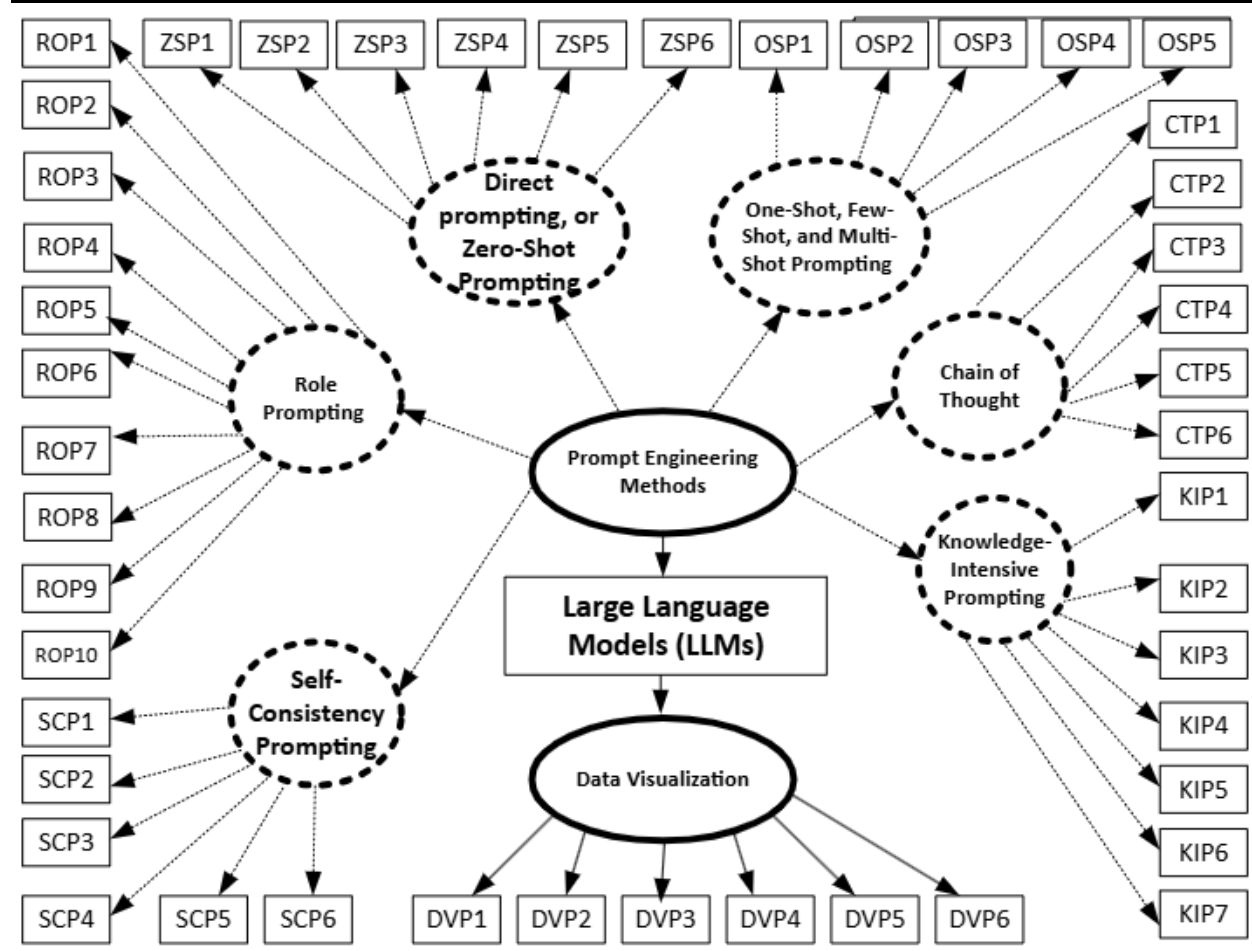


Figure 3: Measurement Frameworks with generated Items from the systematic review

Data Visualization Measurement Instrument

Table 2 presents the measurement items for data visualization, including descriptions of the six items and their relationships.

Table 2. Data Visualization Measurement Items

Coded Items	Items and Variable
DVP	<i>Data Visualization</i>
DVP1	Prompt engineering enables a data visualization system to display a visual design that highlights areas and values.
DVP2	Prompt engineering enables data visualization to show interactivity.
DVP3	Prompt engineering enables data visualization to deliver information in different formats.
DVP4	Prompt engineering enables data visualization to support real-time question-answering.
DVP5	Prompt engineering enables data visualization to use alerting features.
DVP6	Prompt engineering enables data visualization to show key performance indicators.

Role Prompting Measurement Instrument

Table 3 presents the measurement items for role prompting, including descriptions of the ten items and their relationships.

Table 3. Role Prompting Measurement Items

Coded Items	Items and Variable
ROP	<i>Role Prompting</i>
ROP1	The prompt instructed the model to answer questions in the context of a role.
ROP2	The prompt allows the model to produce answers that match my expectations.
ROP3	The prompt allows the model to enhance its response
ROP4	The prompt is understandable and clear, allowing the model to set clear objectives.
ROP5	The prompt allows the model to produce answers that match specific roles.
ROP6	The prompt allows the model to produce task-specific identities.
ROP7	The prompt allows the model to adapt its focus and improve its output as user input changes.
ROP8	The prompt allows the model to understand a follow-up question.
ROP9	The prompt directs the model to concentrate on parts of the previous answer,
ROP10	The prompt allows the model to refine its response in real time.

Direct Prompting, also known as Zero-shot Prompting

Table 4 presents the measurement items for direct prompting, also known as zero-shot prompting, along with descriptions of the six items and their relationships.

Table 4. Zero-Shot Prompting Measurement Items

Coded Items	Items and Variable
ZSP	Direct Prompting or Zero-Shot Prompting
ZSP1	The prompt allows the model to perform tasks based on general instructions without specific training data.
ZSP2	The prompt provides the model with instructions without examples.
ZSP3	The prompt allows the model to enhance its responses by fine-tuning the instructions
ZSP4	The prompt allows the model to enhance its responses by reinforcement learning via human feedback
ZSP5	The prompt allows the model to execute a direct task
ZSP6	The prompt relies entirely on the model's pre-trained knowledge to fulfill the request.

One-Shot, Few-Shot, and Multi-Shot Prompting

Table 5 presents the measurement items for one-shot, few-shot, and multi-shot prompting, along with descriptions of the five items and their relationships.

Table 5. One-Shot, Few-Shot, and Multi-Shot Prompting Measurement Items

Coded Items	Items and Variable
OSP	One-shot, few-shot, and multi-shot prompting
OSP1	The prompt instructs the model to learn a new task with a few examples.
OSP2	The prompt demonstrates that the model can understand complex tasks
OSP3	The prompt demonstrates that the model can understand high-quality examples.
OSP4	The prompt shows that the model can delegate tasks, thereby improving its performance.
OSP5	The prompt enables the model to perform new tasks with little training data, utilizing existing knowledge to give accurate answers

Chain of Thought Prompting

Table 6 presents the measurement items for chain-of-thought prompting, along with descriptions of the six items and their relationships.

Table 6. Chain-of-Thought Prompting Measurement Items

Coded Items	Items and Variable
-------------	--------------------

CTP	Chain-of-thought prompting
CTP1	The prompt provides intermediate reasoning steps to guide a model's responses.
CTP2	The prompt offers a clear structure for the model's reasoning process.
CTP3	The prompt improves the model's reasoning by breaking down tasks into step-by-step processes.
CTP4	The prompt shows that the model has accuracy in complex problem-solving.
CTP5	The prompt allows the model to automatically generate diverse reasoning chains, improving accuracy and robustness.
CTP6	The prompt makes it easier for users to understand how the model reaches its conclusions.

Self-Consistency Prompting

Table 7 presents the measurement items for self-consistency prompting, along with descriptions of the six items and their relationships.

Table 7. Self-Consistency Prompting Measurement Items

Coded Items	Items and Variable
SCP	Self-consistency prompting
SCP1	The prompt shows consistency in the responses provided by the models.
SCP2	The prompt shows efficiency in fields that require high precision and proof.
SCP3	The prompt shows that the model generates multiple reasoning chains and selects the most consistent final answer.
SCP4	The prompt shows that the model can enhance accuracy through diverse approaches.
SCP5	The prompt allows the model to search and access external sources relevant to the prompt.
SCP6	The prompt shows that the model can get new information outside the LLM's original training dataset for consistency checking

Knowledge-Intensive Prompting

Table 8 presents the measurement items for knowledge-intensive prompting, along with descriptions of the seven items and their relationships.

Table 8. Knowledge-Intensive Prompting Measurement Items

Coded Items	Items and Variable
KIP	Knowledge-intensive prompting
KIP1	The prompt helps the model to understand critical knowledge.
KIP2	The prompt allows the model to generate new knowledge.
KIP3	The prompt allows the model to make accurate predictions.
KIP4	The prompt encourages the model to produce outputs and predictions based on factual information.
KIP5	The prompt leverages the model's ability to generate potentially useful information before generating a final response.
KIP6	The prompt improves the model's commonsense reasoning without requiring fine-tuning.
KIP7	The prompt shows that the model uses formal logic and reasoning frameworks to solve problems.

Discussion

The results indicate that AI prompt engineering enables LLMs to generate high-quality visualizations when consistent measurement items are used to assess them. Moreover, the researchers conducted a systematic literature review and identified 40 items measuring successful AI prompt engineering and 6 measuring data visualization.

Role prompting enables data visualization by assigning specific personas to AI. That means the role of prompting is to instruct LLMs to act as experts or specialists. The LLMs will perform data visualization as experts in data analytics. In this sense, data visualization can be tailored to the context and tone to produce precise visualizations rather than vague ones. In this scenario, the visualizations are better aligned with business objectives and meet business requirements.

Direct prompting (zero-shot prompting) facilitates data visualization by enabling LLMs to produce precise visuals, such as histograms, bar charts, pie charts, and tree maps, using natural-language instructions, recommendations, and directives. All accurate visualizations are created solely through direct prompts to LLMs, without the need for clarification requests or additional example interactions.

One-shot, few-shot, and multi-shot prompting methods enable accurate data visualization, such as a scatter plot, by including examples (shots) in the prompt to specify the target output patterns. These techniques minimize the need for extensive training data, improve accuracy, and help organize complex requests to ensure consistent visualization.

Chain-of-thought prompting facilitates data visualization by decomposing complex analysis tasks into clear, sequential phases—such as defining data structures, cleaning data, and selecting chart types that fit the problem and align with user intuition and judgment. This systematic method helps ensure translation from natural language to visualization techniques, enhancing predictability, precision, and interpretability.

Self-consistency prompting enhances data visualization coherence by generating multiple varied reasoning paths for a prompt and selecting the most accurate and suitable final answer. This approach minimizes errors by relying on reasoning processes and identifying the most common and reliable visualizations.

Knowledge-intensive prompts enhance data visualization by integrating structured external knowledge with LLMs. This approach automatically generates, analyzes, and refines visual outputs, converting unstructured data into structured formats such as triplets. This structured data can then be mapped to visual elements such as graphs, charts, and heat maps, enabling deeper insights and more complex predictive and prescriptive analytics.

We note that current prompt engineering classifications primarily describe techniques such as zero-shot, few-shot, chain-of-thought, role prompting, self-consistency, and knowledge-intensive prompting as separate approaches to engaging with LLMs. Conversely, our framework integrates these techniques into AI-driven data visualization and organizes them into a measurement framework that delineates measurable dimensions and indicators. Instead of merely listing prompting methods, our framework offers a theoretically grounded foundation for applying prompt engineering in empirical studies and exploring its impact on data visualization results.

Conclusion

This paper investigated how AI prompt engineering enables data visualization. Because of their versatility and general-purpose nature, LLMs can handle a variety of tasks (Chiarello et al., 2024). Thus, organizations have asked AI-based LLMs for help with data visualization because they have predictive power. In this sense, they have learned to detect patterns, relationships, examples, feedback, and context in prompt engineering. This research uses a systematic literature review to develop measurement tools for data visualization and for role prompting, direct prompting (also known as zero-shot), one-shot, few-shot, and multi-shot prompting, chain-of-thought prompting, self-consistency prompting, and knowledge-intensive prompting.

Therefore, a framework was developed combining structural and measurement research models. The framework allowed us to identify the study's boundaries, thereby enabling a systematic literature review and content analysis, as well as the SPAR-4-SLR methodology by Garg & Sharma (2026).

From the initial pool of 211 returned articles, 99 that did not meet the inclusion criteria were excluded, along with 26 that focused solely on AI or lacked prompt engineering and data visualization. Ultimately, 86 articles were analyzed in detail. Our findings include six indicators assessing the latent variable of data visualization; ten for role prompting; six for direct (zero-shot) prompting; five for one-shot, few-shot, and multi-shot prompting; six for chain-of-thought prompting; six for self-consistency prompting; and seven for knowledge-intensive prompting. Both theoretical and practical insights are offered.

Theoretical Contributions

This study demonstrates a theoretical relationship between prompt engineering and HCI, and between LLMs and HCI. Of course, AI is connected to existing skills like decision-making, critical thinking, problem-solving, and written and oral communication. But prompt engineering skills are specific to AI; in addition to the existing ones, they all form the backbone of skills developed using AI applications. This research is among the first to propose measurement tools to assess the effectiveness of prompt engineering in enabling accurate and consistent data visualization. Currently, data analytics software such as Tableau and Power BI is well-suited for data visualization. However, this existing software requires human involvement because the analysis typically demands a knowledge-intensive approach to achieve comprehensive data visualization. Instead of using, for instance, Tableau, the users can use prompt engineering and LLMs to enable data visualization. This research extends the theoretical background of prompt engineering to data visualization. This research added new measurement tools to the existing list of AI and data visualization measures in the literature. It also used systematic literature review methods based on the SPAR-4-SLR methodology of Garg & Sharma (2026) to develop measures.

Practical Contributions

This study proposes that newly developed measurement tools for prompt engineering should include items and indicators that cover role prompting; direct prompting (also known as zero-shot), one-shot, few-shot, and multi-shot prompting; chain-of-thought prompting; self-consistency prompting; and knowledge-intensive prompting. Most users must learn existing data visualization software before creating visualizations. Empirical research indicates that LLMs can produce visualizations from natural-language prompts (Li et al., 2025); however, these studies do not suggest that prompt engineering can replace dedicated visualization tools. Instead, LLMs can support users in creating visualizations, especially those with limited visualization experience. In other terms, their natural language and the use of LLMs can generate accurate data visualization. Moreover, the challenge of developing prompt engineering that is understandable to LLMs while avoiding algorithmic bias is mitigated by the measurement tools proposed in this research. Furthermore, this research would help users and organizations determine which prompt-engineering methods to use to meet data visualization requirements.

Future Research and Limitations

Future research should expand on this work using quantitative methods to test the validity and reliability of the items generated in this study, including convergent and discriminant validity. The items or indicators created herein should be included in a survey to collect data for testing the measurement model using SEM and SmartPLS. A limitation of this study is that the measurement tools were developed through a systematic literature review. It would be fruitful to include other qualitative analyses, such as interviews and expert evaluation, in the methodology. In addition, the systematic literature review can be strengthened using a validity and reliability study.

The identified measurement items belong to an initial conceptual framework derived from a systematic literature review, rather than a finalized measurement scale. Future researchers can modify these items into questionnaire statements, validate content through expert reviews, pilot test them, and assess construct validity and reliability via methods such as exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and reliability testing. Additionally, we note how the framework can be further refined and validated across different application domains, AI platforms, and user groups. These clarifications clarify the framework's purpose and provide practical guidance for further empirical studies.

We recognize that ongoing advancements in prompting techniques, AI models, and data visualization tools may lead to new constructs or alter the significance of the measurement items discussed here. Therefore, this framework should be seen as a starting point that needs regular updates and empirical checks as generative AI continues to develop. We also recommend future research to update and validate the framework with new studies and emerging AI platforms.

References

- Akyildiz, T. S. (2026). Prompt engineering and creative pedagogical thinking: Insights from a qualitative study with teacher candidates. *Thinking Skills and Creativity*, 60.
<https://doi.org/10.1016/j.tsc.2025.102105>.
- Chaichana, T., Reeve, G., Jaisan C., & Chakrabandhu, Y. (2024). Modeling and assessing new SME digital business status for visualizing virtual economics and sustainability economic indicators: Empirical evidence from the poultry business. *Heliyon*. 10.
<https://doi.org/10.1016/j.heliyon.2024.e30624>
- Chen, B., Zhang, Z., Langrene, N., & Zhu, S. (2026). Unleashing the potential of prompt engineering for large language models. *Patterns Review*. <https://doi.org/10.1016/j.patter.2025.101260>
- Chiarello, F., Giordano, V., Spada, I., Barandoni, S., & Fantoni, G. (2024). Future applications of generative large language models: A data-driven case study on ChatGPT. *Technovation*. 133.
<https://doi.org/10.1016/j.technovation.2024.103002>.
- Chonkker, M. S., Verma, D., & Kar, A. K. (2017). Review of technology adoption framework in mobile commerce. Information Technology and Quantitative Management. *Procedia Computer Science*. 122, 888-895. <https://doi.org/10.1016/j.procs.2017.11.451>
- Du, M., Amor, R., Ma, K-L., & Wünsche, B-C. (2026). Data visualization for improving financial literacy: A systematic review. *Visual Informatics*. Vol. 10.
<https://doi.org/10.1016/j.visinf.2025.100272>
- Endler, M., Rahmani, N., & Khan, F. (2026). Analysis of the AI prompt used to improve safety management system accessibility. *Journal of Loss Prevention in the Process Industries*. 99.
<https://doi.org/10.1016/j.jlp.2025.105834>
- Garg, S. & Sharma, A. K. (2026). Business implications of climate change: A systematic literature review adopting the ADO-TCM framework. *Journal of International Money and Finance*. 163.
<https://doi.org/10.1016/j.jimonfin.2026.103549>

- Garg, A., Soodhani, N., & Rajendran, R. (2025). Enhancing data analysis and programming skills through structured prompt training: The impact of generative AI in engineering education. *Computers and Education: Artificial Intelligence*. 8. <https://doi.org/10.1016/j.caeai.2025.100380>
- Garry, M., Chan, W. M., Foster, J., & Henkel, L. A. (2024). Large language models (LLMs) and the institutionalization of misinformation. *Trends in Cognitive Sciences*. 28, 12. <https://doi.org/10.1016/j.tics.2024.08.007>
- Hwang, M., Lee, K.-H., & Lee, H.-K. (2025). A word to the wise: Crafting impactful prompts for ChatGPT. *System*. 133. <https://doi.org/10.1016/j.system.2025.103756>
- Hendavi, S., Kanan, T., Elbes, M., Mughaid, A., & Alzubi, S. (2026). Automated prompt engineering pipelines: fine-tuning LLMs for enhanced response accuracy. *Expert Systems With Applications*. 298. <https://doi.org/10.1016/j.eswa.2025.129689>
- Ilagana, J. R., Ilagana, J. B., Basalloa, C. L. & Alabastro, Z. M. (2024). Exploratory prompting of large language models to act as co-pilots for augmenting business process work in document classification. *Procedia Computer Science*. 237, 420–425. 10.1016/j.procs.2024.05.123.
- Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*. 6. <https://doi.org/10.1016/j.caeai.2024.100225>
- Koralewska, M., Dratwinskia, F., & Goleca, P. (2025). Usability of LLM-driven loss function in Automatic Prompt Engineering for e-commerce. *Procedia Computer Science*. 270, 1631–1640. 10.1016/j.procs.2025.09.283
- Lawson-Body, A., Jensen, J., & Lawson-Body, L. (2024). Big Data Usage, Natural Language Processing, and Ethics in Developing Countries: Instrument Development Using Systematic Literature Review, *Issues in Information Systems*, 25(2), 379-396.
- Li, Y., Liu, P., Wang, H., Chu, J., & Wong, W. E. (2025). Evaluating large language models for software testing. *Computer Standards & Interfaces*. 93. <https://doi.org/10.1016/j.csi.2024.103942>
- Lian, Y., Hao, J., Zeng, W. & Luo, Q. (2025). A survey of visual insight mining: Connecting data and insights via visualization. *Visual Informatics*. 9. <https://doi.org/10.1016/j.visinf.2025.100271>
- Mosen, J., Gerbl, J., Williams, P., & Schubert, P. (2026). Work Trace Visualizer: Connecting and Visualizing Digital Trace Data to Understand Workgroups and Collaborative Work. *Procedia Computer Science*. 278, 637–647.
- Palvia, P., YK, C. P., Kakhki, M. D., Ghoshal, T., Uppala, V., & Wang, W. 2017. "A Decade Plus Long Introspection of Research Published in Information & Management," *Information & Management* (54:2), pp. 218-227.
- Petzel, Z. W., & Sowerby, L. (2025). Prejudiced interactions with large language models (LLMs) reduce trustworthiness and behavioral intentions among members of stigmatized groups. *Computers in Human Behavior*. 165. <https://doi.org/10.1016/j.chb.2025.108563>
- Phillips-Wren, G., & Hakansson, A. (2025). Towards Using Prompt Engineering in Large Language Models to Assist Decision Making. *Procedia Computer Science*. 270, 5225–5238. 10.1016/j.procs.2025.09.650

- Reisig, M., & Lenz, J. (2026). Application and Evaluation of Communication Protocols in Industry 4.0: Use Case Study with End-to-End Implementation from Data Acquisition to Data Visualization. *Procedia CIRP*. 139, 219–224. 10.1016/j.procir.2025.09.049
- Robertson, J., Ferreira, C., Botha, E., & Oosthuizen, K. (2024). Game changers: A generative AI prompt protocol to enhance human-AI knowledge co-construction. *Business Horizons*. 67, 499-510. <https://doi.org/10.1016/j.bushor.2024.04.008>
- Siddiqui, A., Zerbato, F., & Schuster, D. (2026). Visualizing repetition in process execution variants from partially ordered event data. *Information Systems*. 138. <https://doi.org/10.1016/j.is.2025.102664>
- Su, B., Wang, Z., Lin, Z., Liao, X., Zhou, Z., Zhu, M. & Chen, W. (2026). G2: A customizable web-based framework for authoring interactive visualizations. *Visual Informatics*. <https://doi.org/10.1016/j.visinf.2026.100307>
- Tani, T., Yläkujalaa, A., Metsoa, L., Sinkkonena, T., & Kärria, T. (2025). Prompt Engineering P2X Business Ecosystem with Generative AI. *Procedia Computer Science*. 256, 20-27. 10.1016/j.procs.2025.02.091
- Xiao, J. J., & Meng, K. (2025). AI and personal finance: evaluating rules of thumb generated by ChatGPT. *International Journal of Bank Marketing*. 43, 9, 1969-1989. <https://doi.org/10.1108/IJBM-03-2025-0227>
- Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, Z., & Zhang, Y. (2024). A survey on large language model (LLM) security and privacy: The Good, The Bad, and The Ugly. *High-Confidence Computing*. 4. <https://doi.org/10.1016/j.hcc.2024.100211>.
- Zhang, R., Li, H.-W., Qian, X.-Y., Jiang, W.-B., & Chen, H.-X. (2025). On large language models safety, security, and privacy: A survey. *Journal of Electronic Science and Technology*. <https://doi.org/10.1016/j.jnlest.2025.100301>