CURRENT RESEARCH JOURNAL OF PHILOLOGICAL SCIENCES (ISSN: 2767-3758)

VOLUME: Vol.07 Issue04 2026
Page: - 67-74
DOI: - 10.37547/philological-crjps-07-04-11

RESEARCH ARTICLE OPEN ACCESS

From Minimal Directives to Layered Instructional Frames: Structural-Compositional Models of Prompts in Human-Ai Discourse

Tulanboev Shokhsuvar

1-year doctoral (PhD) student of Andijan State Institute of Foreign Languages, Uzbekistan

Received: 20 March 2026 **Accepted:** 12 April 2026 **Published:** 30 April 2026

ABSTRACT

Prompts have become a primary linguistic interface through which users formulate intentions, specify tasks, and regulate the outputs of large language models. Existing research has generated extensive taxonomies of prompting techniques, yet it often classifies prompts by computational strategy or performance effect rather than by their internal linguistic architecture. This study develops a structural-compositional account of prompts as instructional macroacts in human-AI discourse. A purposive integrative review of 22 foundational and contemporary sources published between 1962 and 2026 was conducted. The sources were coded at three levels: macrostructural organization, semantic-pragmatic composition, and surface realization. The synthesis identifies ten recurrent functional slots: role, goal, task, input/object, context, audience, constraint, demonstration, output format, and verification. These slots combine into seven major structural models ranging from the minimal directive model to task-object, goal-conditioned, constraint-augmented, role-contextualized, demonstration-guided, hierarchical-decomposed, and iterative-dialogic configurations. The findings show that prompt complexity cannot be reduced to word count. It is determined by slot breadth, relational depth, explicitness, and extension across conversational turns. The proposed Prompt Structural-Composition Framework distinguishes the communicative content of a prompt from the linguistic relations that organize it and provides a common analytic basis for comparing prompts across languages. The model is particularly applicable to future English-Uzbek studies because it allows functionally equivalent components to be compared even when their morphological and syntactic realization differs.

Keywords: Prompt; structural-compositional model; instructional discourse; human-AI communication; directive; constraint; prompt architecture; comparative linguistics.

INTRODUCTION

The expansion of generative language-model interfaces has turned prompting into a recurrent form of ordinary linguistic action. A user no longer selects commands exclusively from a predetermined menu; instead, the user formulates a natural-language instruction that may define a goal, identify an object, provide context, impose restrictions, and prescribe the form of the expected response. The prompt therefore functions simultaneously as an utterance, a task specification, and an instrument for regulating an algorithmically generated text. From a linguistic perspective, this makes prompting a distinctive instance of instructional discourse rather than merely a technical input procedure.

The pragmatic basis of such discourse is well established in speech-act theory. Directives are utterances through which a speaker attempts to influence a future action of an addressee (Austin, 1962; Searle, 1969, 1976). Procedural and instructional texts expand this basic directive relation

by arranging actions, conditions, stages, and expected outcomes into a larger communicative unit (Longacre, 1996; Farkas, 1999). In human-AI communication, the addressee is algorithmic rather than human, but the user still encodes an intended operation through linguistic choices. The response is shaped not only by the lexical verb that names the task but also by how the entire prompt distributes background knowledge, requirements, examples, and evaluative criteria.

Prompt research has produced several influential classifications. Early work described prompt programming as a way of steering large language models through natural-language instructions and examples (Reynolds & McDonell, 2021). Later studies catalogued reusable prompt patterns (White et al., 2023), classified levels of task detail (Karmaker & Feng, 2023), surveyed prompting techniques (Liu et al., 2023; Schulhoff et al., 2024), and proposed practitioner-oriented frameworks such as the Prompt Canvas (Hewing & Leinhos, 2024). Recent work has moved toward explicitly linguistic analysis. PromptPrism, for example, distinguishes functional structure, semantic components, and syntactic patterns (Jeoung et al., 2026). These contributions demonstrate that prompt form matters, but they do not eliminate the need for a compact structural model that explains how functional elements are assembled into complete instructional units.

A central conceptual problem is that a prompting technique is not identical to a prompt structure. Few-shot prompting, role prompting, chain-of-thought prompting, and decomposition describe strategies that may occupy different positions within the same prompt. A single block-structured prompt may contain a role assignment, task directive, two demonstrations, several constraints, and a verification request. Treating each strategy as a mutually exclusive prompt type conceals the internal composition of the utterance. Conversely, measuring prompt quality by length alone obscures the difference between useful specification and redundant expansion. Experimental studies show that explicit input-output requirements, edge-case handling, and stepwise breakdowns can improve task control, while poorly articulated requirements and fragile examples may produce unstable outcomes (Lu et al., 2023; Ma et al., 2025; Zi et al., 2025).

The present study addresses this gap by conceptualizing the prompt as a compositional instructional frame. It asks three questions: (1) Which semantic-pragmatic components recur in prompts addressed to language models? (2) Through which macrostructural models are these components organized? (3) How can structural complexity be described without equating it with verbosity? The study contributes a Prompt Structural-Composition Framework (PSCF) consisting of ten functional slots and seven recurrent macrostructural models. Its purpose is analytical rather than prescriptive: it offers categories for describing naturally occurring prompts, comparing prompt designs, and examining equivalent functions across languages.

METHODS

1. Research design and source selection. The study employed a theory-building integrative review combined with qualitative linguistic synthesis. This design was selected because the aim was not to estimate the mean effectiveness of a prompting technique but to reconstruct the internal architecture shared by heterogeneous prompts. The analytical corpus comprised 22 sources. Six established the pragmatic, functional, or procedural foundations of instruction; sixteen addressed prompt learning, instruction following, prompt patterns, constraints, decomposition, user practices, requirement articulation, specificity, or prompt taxonomies. Publication dates ranged from 1962 to 2026, while the prompt-specific subset was concentrated between 2021 and 2026.

Sources were included when they met at least one of four criteria: they explicitly defined the components of a prompt; proposed a prompt taxonomy or reusable pattern; experimentally manipulated task detail, constraints, examples, or decomposition; or documented how users construct and revise prompts. Studies concerned exclusively with continuous or soft prompts, parameter tuning, or model-internal optimization were excluded because their prompt representations are not naturally occurring linguistic texts. Task-specific studies were retained only when their analysis identified general structural information such as input specifications, output requirements, constraints, examples, or stepwise organization.

The review was purposive rather than exhaustive. Its validity rests on conceptual coverage and triangulation across linguistics, natural-language processing, human-computer interaction, and software engineering. The corpus included major syntheses (Liu et al., 2023; Schulhoff et al., 2024), explicit structural frameworks (White et al., 2023; Karmaker & Feng, 2023; Hewing &

Leinhos, 2024; Jeoung et al., 2026), experimental studies of constraints and specificity (Lu et al., 2023; Ma et al., 2025; Zi et al., 2025), and studies of prompt construction in practice (Zamfirescu-Pereira et al., 2023; Liang et al., 2025).

2. Coding procedure. Each source was examined at three analytical levels. The first was macrostructure: whether a prompt was organized as a single clause, a sequence of directives, a set of labelled blocks, a hierarchy of subtasks, or a multi-turn exchange. The second was semantic-pragmatic composition: which communicative functions were encoded, including the assignment of a role, statement of purpose, task, input, context, audience, constraints, demonstrations, output form, and verification criteria. The third was surface realization: the grammatical and graphic devices used to signal these functions, such as imperatives, declaratives, interrogatives, headings, bullet points, delimiters, numbered stages, examples, and anaphoric references.

Coding proceeded deductively from speech-act theory and functional grammar and inductively from recurrent components named in the prompt literature. Functionally similar labels were merged. For example, persona, expert identity, and professional stance were grouped under role; desired response shape, schema, and presentation type were grouped under output format; evaluation rubric, self-check, and acceptance test were grouped under verification. A component was treated as a functional slot only when it could be omitted, expanded, reordered, or expressed by more than one grammatical form without changing the identity of the remaining components.

The emerging model was cross-checked against three external organizing schemes. TELeR was used to test whether the framework captured degrees of task specification; the Prompt Canvas was used to check coverage of practitioner-visible elements; and PromptPrism was used to verify that macrostructure, semantic content, and syntactic realization remained analytically distinct. The resulting framework is a conceptual synthesis. It does not claim that one model is universally superior, and no model-performance experiment was conducted.

RESULTS

1. The ten-slot compositional frame. The synthesis indicates that a prototypical prompt can be represented as an ordered but flexible set of ten functional slots: P = <R, G, T, I, K, A, C, D, F, V>. Here R denotes role, G goal, T task, I input or object, K context, A audience, C constraints, D demonstrations, F output format, and V verification. The task slot is the only structurally indispensable component in a prototypical instructional prompt; all other slots are optional in principle but may become functionally necessary when the task is ambiguous, specialized, or tightly regulated. The brackets do not imply a fixed sequence. The slots may be reordered, embedded, repeated, or distributed across turns.

The distinction between goal and task is particularly important. A goal specifies why the output is needed or what communicative effect it should achieve, whereas a task identifies the operation to be performed. In the prompt 'Prepare a concise briefing so that first-year students can understand the policy; summarize the report in 180 words,' the first clause encodes goal and audience, while the second encodes task, input, and quantitative constraint. Likewise, context differs from input. Input is the specific material to be transformed or analyzed; context supplies background assumptions necessary for interpreting that material.

Constraints and verification form separate control layers. A constraint limits admissible procedures or outputs before generation, as in 'Do not add information not present in the source.' Verification requires an explicit post-generation check, as in 'Confirm that every statistic in the answer appears in the source.' The distinction matters because a prompt may prohibit an action without requesting an audit, or it may request an audit even when no explicit prohibition is stated. Demonstrations are also distinct from context: they model a relation between input and output rather than merely supplying background knowledge.

Table 1. Functional slots in the Prompt Structural-Composition Framework

Code	Slot	Communicative function	Typical realization
R	Role	Assigns an expert, institutional, or interactional position	Act as an editor; You are advising...

Code	Slot	Communicative function	Typical realization
G	Goal	States the intended purpose or communicative effect	The aim is to inform; so that beginners can...
T	Task	Names the operation to be performed	Summarize; compare; classify; explain
I	Input/object	Identifies the material or referent acted upon	the following passage; the attached table
K	Context	Supplies background, situation, or assumptions	This report concerns...; In the current case...
A	Audience	Specifies the intended recipient of the output	for specialists; for first-year students
C	Constraint	Limits content, procedure, style, length, or sources	do not invent; use only; under 200 words
D	Demonstration	Models expected input-output behaviour	Example 1: Input... Output...
F	Output format	Defines external organization or data structure	one paragraph; a table; JSON; numbered steps
V	Verification	Requests checking, testing, or criterion-based revision	verify; cross-check; revise until all criteria are met

2. Seven structural-compositional models. The ten slots do not occur randomly. They combine into recurrent macrostructural configurations that differ in informational breadth and organizational depth. Seven models emerged from the synthesis. These models form a continuum rather than a hierarchy of quality: a minimal prompt may be fully adequate for a familiar, low-risk task, whereas a layered model is more appropriate when the task contains several constraints, specialized inputs, or evaluation conditions.

Model 1, the minimal directive, consists of a task alone: [T], for example, 'Summarize.' Its interpretation depends almost completely on the surrounding conversation or visible input. Model 2, the task-object model, combines operation and target: [T + I], as in 'Summarize the article.' This is the basic transitive pattern of prompting. Model 3, the goal-conditioned model, adds an explicit communicative purpose: [G + T + I], enabling the same task to be shaped by a desired outcome. Model 4, the constraint-augmented model, adds one or more limits and usually an output specification: [T + I + C + F]. This is the dominant architecture of controlled single-turn prompting.

Model 5, the role-contextualized model, frames the task through an assigned perspective and situational background: [R + K + G + T + I + C/F]. Role does not replace task specification; it narrows the interpretive stance from which the task is performed. Model 6, the demonstration-guided model, adds one or more input-output examples: [K/R + T + I + D + C + F]. It is structurally distinctive because part of the prompt is iconic rather than purely directive: the example shows the expected relation instead of only describing it. Few-shot and chain-of-thought demonstrations are realizations of this model (Wei et al., 2022).

Model 7, the hierarchical-decomposed model, distributes the instruction across labelled blocks and ordered subtasks. Its canonical organization is [R/K] -> [G/T] -> [I] -> [C] -> [F] -> [V], with the task itself divided into T1, T2, ... Tn. Decomposed prompting demonstrates the computational value of modular subtask organization (Khot et al., 2023), but linguistically the key property is hierarchical dependency: each local directive contributes to a superordinate macrogoal. A dialogic extension of this model occurs when output evaluation produces a follow-

up prompt, P1 -> R1 -> P2 -> R2. The second prompt may add a missing slot, strengthen a constraint, or repair reference, making prompt composition extend across conversational turns rather than remain inside a single message.

Table 2. Prompt structural-composition models and principal communicative affordances

No.	Model	Core formula	Principal communicative affordance
1	Minimal directive	T	Fast action where the object and context are already recoverable
2	Task-object	T + I	Explicitly links an operation to its target
3	Goal-conditioned	G + T + I	Aligns the operation with a communicative purpose
4	Constraint-augmented	T + I + C + F	Controls admissible content and response shape
5	Role-contextualized	R + K + G + T + I	Frames interpretation through role and situation
6	Demonstration-guided	T + I + D + C + F	Shows rather than only describes the expected mapping
7	Hierarchical-decomposed / dialogic	T1...Tn + C + F + V; P1-R1-P2-R2	Coordinates subtasks, checkpoints, and iterative repair

3. Relations among components. Structural composition is determined not only by which slots are present but by the relations among them. Five relations are recurrent. Accumulation places components in a linear sequence, as in task + constraint + format. Coordination joins directives of equal rank: 'Identify the claims, classify the evidence, and summarize the conclusion.' Embedding places one component inside another, for example when audience is encoded within a style constraint: 'Explain the concept in language accessible to twelve-year-old learners.' Hierarchy arranges directives as subordinate stages under one macrogoal. Recurrence distributes components across turns, where a later prompt re-specifies or repairs an earlier instruction.

Graphic organization performs a genuine discourse function. Headings such as ROLE, TASK, CONTEXT, CONSTRAINTS, and OUTPUT FORMAT mark boundaries between semantic blocks; bullet points coordinate requirements; numbering signals procedural order; quotation marks and XML-like tags delimit source material from instructions. These devices are not merely visual decoration. They externalize relations that would otherwise depend on syntax and inference. Current prompt guidance frequently recommends explicit structuring and delimiters, while PromptPrism demonstrates that delimiter changes and semantic reordering can be analyzed as independent dimensions of prompt sensitivity (Jeoung et al., 2026).

A prototypical block-structured prompt may therefore be represented as follows:

ROLE: Act as an academic editor.

GOAL: Produce a publication-ready abstract for an interdisciplinary audience.

TASK AND INPUT: Rewrite the abstract supplied below.

CONSTRAINTS: Preserve the findings and citations; add no new claims; use 180-200 words.

OUTPUT FORMAT: One paragraph followed by five keywords.

VERIFICATION: Check the word limit and confirm that every numerical result is retained.

The example contains six explicit blocks but more than six semantic slots: audience is embedded in the goal, input is integrated with the task, and two separate content restrictions occupy the constraint block. Consequently, the number of labelled sections should not be treated as identical to the number of functional components.

The synthesis also shows that complexity is not equivalent to length. Structural complexity has three dimensions: slot breadth, or the number of distinct functions encoded; relational depth, or the extent to which components are embedded or hierarchically ordered; and turn extension, or the distribution of instruction across a dialogue. A long prompt may be structurally simple if it consists mainly of background context, while a short prompt may be compositionally dense if it combines task, quantitative restriction, source limitation, format, and verification in one sentence.

DISCUSSION

The proposed framework reinterprets prompt design as a problem of discourse composition. Existing surveys are indispensable for identifying prompting techniques, but a technique-oriented taxonomy and a linguistic architecture answer different questions. Chain-of-thought specifies a reasoning demonstration or procedure; role prompting fills the role slot; few-shot prompting fills the demonstration slot; output schemas fill the format slot; and self-critique or acceptance tests fill the verification slot. These devices can coexist within a single prompt. The PSCF therefore treats them as compositional resources rather than mutually exclusive prompt genres.

This distinction clarifies the relationship between the present model and previous frameworks. White et al. (2023) describe reusable prompt patterns for solving recurrent interactional problems. TELeR organizes prompts according to the amount and type of task detail supplied (Karmaker & Feng, 2023). The Prompt Canvas foregrounds practical elements including persona, audience, context, output, references, and stepwise procedure (Hewing & Leinhos, 2024). PromptPrism adds a linguistically inspired hierarchy of functional, semantic, and syntactic analysis (Jeoung et al., 2026). The PSCF complements these approaches by concentrating on the internal organization of the instructional macroact: what slots are present, how they are related, and which macrostructure they collectively form.

The results support a requirement-centered rather than ornament-centered view of prompting. Ma et al. (2025) found that training users to articulate clear, complete requirements improved prompt and output quality more effectively than instruction focused on conventional prompting tricks. Zi et al. (2025) likewise identify explicit input-output specifications, edge-case handling, and stepwise breakdown as important contributors to prompt specificity in code generation. These findings align with the constraint-augmented and hierarchical models. A role label or a request to 'think step by step' cannot compensate for an unstated object, missing output condition, or contradictory restriction.

At the same time, adding components indiscriminately is not a universal solution. Prompts may become over-specified, internally inconsistent, or burdened by examples that model the wrong relation. Lu et al. (2023) show that structural and stylistic constraints can reveal persistent generative failures. User studies also show that non-experts often revise prompts opportunistically, overfit them to a small set of examples, and lack stable mental models of how prompt changes affect outputs (Zamfirescu-Pereira et al., 2023; Liang et al., 2025). The relevant analytical question is therefore not 'How long is the prompt?' but 'Which communicative requirement is each component performing, and how is that requirement linked to the others?'

The framework has direct implications for comparative linguistics. Functional equivalence can be established at the slot level even when languages use different grammatical means. English 'Summarize the passage in no more than 120 words and preserve the citations' and Uzbek 'Matnni 120 so'zdan oshirmasdan qisqartiring va manba ishoralarini saqlang' instantiate the same task + input + quantitative constraint + preservation constraint configuration. English relies on a prepositional limit expression and coordinated imperatives; Uzbek uses the agglutinative form -dan oshirmasdan and the polite imperative -ing. A comparative study can thus distinguish structural isomorphism at the functional level from allomorphism at the morphosyntactic level.

The same principle applies to block order and explicitness. English prompts may favor preposed role and context clauses, fixed word order, modal constructions, and prepositional format phrases. Uzbek prompts may encode relations through case, possession, converbs, postpositions, modal predicates, and imperative

morphology. These are hypotheses for empirical verification, not findings of the present review. A balanced English-Uzbek corpus should therefore annotate both the ten functional slots and their language-specific realizations, allowing the researcher to compare component frequency, ordering, implicitness, and grammatical coding without using literal translation as the sole tertium comparationis.

The study has three limitations. First, the review is purposive and concept-oriented rather than a PRISMA-style systematic review. Second, the framework has not yet been tested through inter-annotator agreement on a bilingual corpus. Third, prompt conventions change rapidly as model interfaces, system instructions, tool use, and structured-output mechanisms evolve. Future research should operationalize the ten slots in an annotation manual, test the seven models on naturally produced English and Uzbek prompts, and examine how structural profiles correlate with semantic-pragmatic alignment between prompts and responses.

CONCLUSION

Prompts in human-AI discourse are not homogeneous strings of instructions. They are compositional macroacts in which task, object, purpose, context, role, audience, restrictions, examples, response format, and verification may be selectively combined. The integrative analysis developed a ten-slot Prompt Structural-Composition Framework and identified seven recurrent macrostructural models, from minimal directives to hierarchical and dialogic configurations.

The principal theoretical conclusion is that prompt complexity must be described through function and relation rather than word count. Slot breadth captures how many communicative functions are encoded; relational depth captures coordination, embedding, and hierarchy; turn extension captures the continuation and repair of instruction across dialogue. This distinction makes it possible to compare prompts that differ in length, genre, language, or technical purpose while retaining a common analytical basis.

For comparative English-Uzbek research, the framework offers a defensible tertium comparationis. Equivalent instructional functions can be aligned first, after which their lexical, morphological, syntactic, graphic, and interactional realizations can be compared. Empirical validation on a balanced bilingual corpus may therefore yield both a general model of prompt architecture and language-specific evidence concerning how human intentions are structured for an algorithmic addressee.

REFERENCES

1. Austin, J. L. (1962). How to do things with words. Clarendon Press.

2. Farkas, D. K. (1999). The logical and rhetorical construction of procedural discourse. Technical Communication, 46(1), 42-54.

3. Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). Halliday's introduction to functional grammar (4th ed.). Routledge.

4. Hewing, M., & Leinhos, V. (2024). The Prompt Canvas: A literature-based practitioner guide for creating effective prompts in large language models. arXiv. https://doi.org/10.48550/arXiv.2412.05127

5. Jeoung, S., Chen, Y., Zhang, Y., Wang, S., Ding, H., & Cheong, L. L. (2026). PromptPrism: A linguistically-inspired taxonomy for prompts. In Findings of the Association for Computational Linguistics: EACL 2026 (pp. 1168-1192). Association for Computational Linguistics. https://doi.org/10.18653/v1/2026.findings-eacl.61

6. Karmaker, S. K., & Feng, D. (2023). TELeR: A general taxonomy of LLM prompts for benchmarking complex tasks. In Findings of the Association for Computational Linguistics: EMNLP 2023 (pp. 14197-14203). Association for Computational Linguistics.

7. Khot, T., Trivedi, H., Finlayson, M., Fu, Y., Richardson, K., Clark, P., & Sabharwal, A. (2023). Decomposed prompting: A modular approach for solving complex tasks. In The Eleventh International Conference on Learning Representations.

8. Liang, J. T., Lin, M., Rao, N., & Myers, B. (2025). Prompts are programs too! Understanding how developers build software containing prompts. Proceedings of the ACM on Software Engineering, 2(FSE), Article FSE072. https://doi.org/10.1145/3729342

9. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 55(9), 1-35. https://doi.org/10.1145/3560815

10. Longacre, R. E. (1996). The grammar of discourse (2nd ed.). Plenum Press.

11. Lu, A., Zhang, H., Zhang, Y., Wang, X., & Yang, D. (2023). Bounding the capabilities of large language models in open text generation with prompt constraints. In Findings of the Association for Computational Linguistics: EACL 2023 (pp. 1982-2008). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-eacl.148

12. Ma, Q., Peng, W., Yang, C., Shen, H., Koedinger, K., & Wu, T. (2025). What should we engineer in prompts? Training humans in requirement-driven LLM use. ACM Transactions on Computer-Human Interaction, 32(4), Article 41, 1-27. https://doi.org/10.1145/3731756

13. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730-27744.

14. Reynolds, L., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-7). Association for Computing Machinery. https://doi.org/10.1145/3411763.3451760

15. Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., Li, Y., Gupta, A., Han, H., Schulhoff, S., et al. (2024). The Prompt Report: A systematic survey of prompting techniques. arXiv. https://doi.org/10.48550/arXiv.2406.06608

16. Searle, J. R. (1969). Speech acts: An essay in the philosophy of language. Cambridge University Press.

17. Searle, J. R. (1976). A classification of illocutionary acts. Language in Society, 5(1), 1-23. https://doi.org/10.1017/S0047404500006837

18. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E. H., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35, 24824-24837.

19. 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. https://doi.org/10.48550/arXiv.2302.11382

20. Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Article 437, pp. 1-21). Association for Computing Machinery. https://doi.org/10.1145/3544548.3581388

21. Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., & Ba, J. (2023). Large language models are human-level prompt engineers. In The Eleventh International Conference on Learning Representations.

22. Zi, Y., Menon, H., & Guha, A. (2025). More than a score: Probing the impact of prompt specificity on LLM code generation. In Proceedings of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (pp. 2380-2402). Association for Computational Linguistics.