Information Design with Large Language Models: An Annotated Reading List

TAO LIN
The Chinese University of Hong Kong, Shenzhen
and
SAFWAN HOSSAIN
Harvard University
and
YILING CHEN
Harvard University

This annotated reading list surveys emerging works at the intersection of information design and large language models (LLMs). While classical information design theory studies signaling in abstract mathematical models, real-world communication often occurs in natural language. We highlight papers that use LLMs to elicit and communicate information through natural-language messages, cover techniques used in "information design + LLM" research such as language-space optimization and LLM proxies, and discuss papers on LLM persuasion. We aim to illustrate the potential of LLMs to bridge the gap between the theory and practice of information design.

Categories and Subject Descriptors: F [**Theory of Computation**]: Theory of computing—*Algorithmic mechanism design*

General Terms: Economics; Design; Human factors; Languages; Theory

Additional Key Words and Phrases: Persuasion, Signaling, Decision-making, Generative AI, LLM

1. INTRODUCTION

Information design asks a deceptively simple question: How should a player strategically select and present information to shape the belief and behavior of others? Traditionally, this question has been studied in stylized models, where the transmitted information is modeled as an abstract random variable (called "signal") correlated with the state of the world. However, real-world communication often occurs in *natural language*, which has long been an unwieldy object for formal analysis. The recent breakthroughs in generative AI, particularly Large Language Models (LLMs), have the potential to bridge this gap.

LLMs offer numerous new opportunities to information design. First, by engaging users in natural conversations, an LLM can elicit users' private preferences that are difficult to elicit through traditional methods (e.g., surveys) and one-shot interactions. Second, LLMs can serve as proxies for human decision-makers who receive information in natural language. Instead of assuming ideal Bayesian decision-makers, we can study how an LLM – prompted to simulate humans or directly make decisions on behalf of humans – responds to linguistic messages, and

Authors' addresses: lintao@cuhk.edu.cn, shossain@g.harvard.edu, yiling@seas.harvard.edu

use that as a testbed for designing communication policies. Third, LLMs offer a way to search over the vast space of natural language and interaction protocols, rather than over the abstract space of information structures (signaling schemes), enabling the design of information strategies that are both theoretically grounded and practically implementable.

This non-exhaustive annotated reading list surveys emerging works at this interface between information design and LLMs. It highlights papers that use generative AI to elicit and communicate information to influence decision-makers (1)(2)(3)(4). Some papers not directly related to information design are included because they provide a toolbox for "information design + LLM" research, such as language optimization techniques (5) and LLM proxies for human decision-makers (6)(7). As information design is closely related to persuasion, some empirical papers on "LLM + persuasion" are also discussed (8)(9)(10). The selected papers aim to give SIGecom readers an entry point to this rapidly developing area, emphasize conceptual connections between classical information design theory and modern generative AI technologies.

(1) "Information Design With Large Language Models" (arXiv, 2025).
Duetting, P., Hossain, S., Lin, T., Leme, R. P., Ravindranath, S. S., Xu, H., and Zuo, S.

This work defines an information design (Bayesian persuasion) problem with framing effect, where the framing of a signal shapes the receiver's prior belief in a non-Bayesian way, while the signal further updates the receiver's belief via Bayes rule. Besides theoretical characterization of the optimal joint design of framing and signaling, the paper uses LLMs to simulate the framing-to-belief mapping of the receiver and to optimize framing using a hill-climbing-based prompt optimization method.

(2) "Verbalized Bayesian Persuasion" (arXiv, 2025).
Li, W., Lin, Y., Wang, X., Jin, B., Zha, H., and Wang, B.

This work formulates Bayesian persuasion as a verbalized game, representing states and signals as texts, and signaling schemes (stochastic mappings from states to signals) as "writing styles" that control LLM's outputs. Equilibrium computation and prompt optimization techniques (e.g, OPRO (5)) are used to jointly optimize the signaling scheme and the receiver's responding strategy in language space, obtaining outcomes that are comparable to the theoretical Bayesian persuasion outcomes.

(3) "AI Realtor: Towards Grounded Persuasive Language Generation for Automated Copywriting" (arXiv, 2025).
Wu, J., Yang, C., Wu, Y., Mahns, S., Wang, C., Zhu, H., Fang, F., and Xu, H.

This work uses LLM to generate textual product descriptions (for houses) to highlight certain features of the product while adhering to the true product attributes. The set of highlighted features is personalized to the buyer, whose preference is elicited by pre-participation survey as well as LLM-assisted inference from revealed preference during interactions. Human-subject experiments demonstrate both attractiveness and validity of LLM-generated descriptions.

(4) "Algorithmic Persuasion Through Simulation" (arXiv, 2023).
Harris, K., Immorlica, N., Lucier, B., and Slivkins, A.

This theoretical work studies a Bayesian persuasion problem where the sender infers the receiver's private type by querying an oracle that simulates the receiver's behavior given signals. Such an oracle can be implemented by LLM (see (6)(7) for justification). The optimal joint querying and signaling strategy is characterized.

(5) "Large Language Models as Optimizers" (ICLR, 2024).
Yang, C., Wang, X., Lu, Y., Liu, H., Le, Q. V., Zhou, D., and Chen, X.

This influential work from Google DeepMind proposes a prompt optimization method, OPRO (Optimization by PROmpting), which treats an LLM as a gradient-free optimizer that, given the problem description in natural language, iteratively improves prompts based on previously tried prompts and their evaluation scores. OPRO and similar methods have been used to optimize linguistic strategies in previous works on information design with LLMs (1)(2).

(6) "Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?" (EC, 2024).
Filippas, A., Horton, J. J., and Manning, B. S.

This work prompts LLMs with different "personas" and performs classical behavioral economics and social science experiments on those LLMs. By comparing LLM and human responses, this work demonstrates LLMs' ability to simulate human populations with diverse preferences, thereby providing new opportunities and drastically reducing the cost of human-subject experiments. Such an LLM proxy idea has been used by existing works on information design with LLM (1)(4).

(7) "LLM-Powered Preference Elicitation in Combinatorial Assignment" (arXiv, 2025).
Soumalias, E., Jiang, Y., Zhu, K., Curry, M., Seuken, S., and Parkes, D. C.

A more recent work using LLMs as proxies for humans, specifically for preference elicitation in combinatorial assignment. Students describe their preferences for course assignments to LLMs in natural language. Assignment algorithms then make queries to LLMs instead of students, alleviating students' burden.

(8) "The Persuasive Power of Large Language Models" (AAAI Conference on Web and Social Media, 2024).
Breum, S. M., Egdal, D. V., Gram Mortensen, V., Møller, A. G., and Aiello, L. M.

Information design is closely related to persuasion. This is a foundational work on the empirical evaluation of the persuasion ability of LLMs. It conducts systematic experiments where human subjects are exposed to arguments generated by either humans or LLMs, and measures how these arguments influence participants' beliefs or decisions. LLM-generated arguments are demonstrated to be comparable and sometimes more persuasive than human-written ones.

(9) "Persuasion with Large Language Models: a Survey" (arXiv, 2024).
Rogiers, A., Noels, S., Buyl, M., and Bie, T. D.

An empirical survey of how LLMs persuade humans in real-world domains such as politics, marketing, public health, and e-commerce. It identifies key factors for persuasive effectiveness, including personalization, interactivity, prompt design, model scale, and disclosure of AI authorship. It also surveys experiment paradigms, ethical considerations, and regulatory landscape.

(10) "The Earth is Flat because...: Investigating LLMs' Belief towards Misinformation via Persuasive Conversation" (ACL, 2024).
Xu, R., Lin, B., Yang, S., Zhang, T., Shi, W., Zhang, T., Fang, Z., Xu, W., and Qiu, H.

Complementing works on LLMs' ability to persuade humans (8)(9), this ACL 2024 Outstanding Paper flips the lens to examine LLMs' susceptibility to being persuaded by misinformation. The authors curated a dataset, "Farm (Fact-to-Misinform)", consisting of factual questions paired with wrong answers and arguments. They presented such misinformation to LLMs in multi-turn dialogues, finding that LLMs' initial correct beliefs can be manipulated over turns.

REFERENCES

Breum, S. M., Egdal, D. V., Gram Mortensen, V., Møller, A. G., and Aiello, L. M. 2024. The Persuasive Power of Large Language Models. *Proceedings of the International AAAI Conference on Web and Social Media 18*, 152–163.

Duetting, P., Hossain, S., Lin, T., Leme, R. P., Ravindranath, S. S., Xu, H., and Zuo, S. 2025. Information Design With Large Language Models. *CoRR abs/2509.25565*.

Filippas, A., Horton, J. J., and Manning, B. S. 2024. Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus? In *Proceedings of the 25th ACM Conference on Economics and Computation*. ACM, New Haven CT USA, 614–615.

Harris, K., Immorlica, N., Lucier, B., and Slivkins, A. 2023. Algorithmic Persuasion Through Simulation. *CoRR abs/2311.18138*.

Li, W., Lin, Y., Wang, X., Jin, B., Zha, H., and Wang, B. 2025. Verbalized Bayesian Persuasion. *CoRR abs/2502.01587*.

Rogiers, A., Noels, S., Buyl, M., and Bie, T. D. 2024. Persuasion with Large Language Models: a Survey. *CoRR abs/2411.06837*.

Soumalias, E., Jiang, Y., Zhu, K., Curry, M., Seuken, S., and Parkes, D. C. 2025. LLM-Powered Preference Elicitation in Combinatorial Assignment. *CoRR abs/2502.10308*.

Wu, J., Yang, C., Wu, Y., Mahns, S., Wang, C., Zhu, H., Fang, F., and Xu, H. 2025. AI Realtor: Towards Grounded Persuasive Language Generation for Automated Copywriting. *CoRR abs/2502.16810*.

Xu, R., Lin, B., Yang, S., Zhang, T., Shi, W., Zhang, T., Fang, Z., Xu, W., and Qiu, H. 2024. The Earth is Flat because...: Investigating LLMs' Belief towards Misinformation via Persuasive Conversation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational*

Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Bangkok, Thailand, 16259–16303.

Yang, C., Wang, X., Lu, Y., Liu, H., Le, Q. V., Zhou, D., and Chen, X. 2024. Large Language Models as Optimizers. In *The Twelfth International Conference on Learning Representations*.