

Quick Answer: How Will Prompt Engineering Impact the Work of Data Scientists?

Published 9 June 2023 - ID G00794952 - 6 min read

Afraz Jaffri

Initiatives: [Analytics, BI and Data Science Solutions](#); [Evolve Technology and Process Capabilities to Support D&A](#)

Generative AI models produce outputs based on natural language inputs or “prompts.” Prompts are generally seen as simple instructions, but data and analytics leaders should utilize data scientists’ expertise to build prompts capable of more complex tasks and accelerate their own productivity.

Additional Perspectives

- [Summary Translation: Quick Answer: How Will Prompt Engineering Impact the Work of Data Scientists?](#)
(11 August 2023)

Quick Answer

How will prompt engineering impact the work of data scientists?

- Prompt engineering can customize large language models (LLMs) as part of the workflow to power generative AI applications. Data scientists’ underlying knowledge of LLM architecture makes them well-suited to this task.
- Prompt engineering will be used with LLMs to accelerate common machine learning tasks on text and images, such as categorization, summarization, annotation and sentiment analysis.
- Code generation enhanced by prompt engineering will accelerate many tasks, including exploratory data analysis, model selection, and training and model evaluation. These tasks will become increasingly declarative as specialized models perform subtasks in reaching an overall objective.
- Adversarial attacks and unwanted bias can be perpetrated through intentional or unintentional prompt manipulation. The tasks of prompt validation and testing are likely to initially be the responsibility of data scientists.

More Detail

Prompt Engineering as Part of the Data Science Workflow

Prompt engineering is the discipline of providing inputs, in the form of text or images, to generative AI models to confine the set of responses the model can produce to a set that produces a desired outcome. Prompt engineering is also referred to as “in-context learning,” where examples are provided to further guide the model.

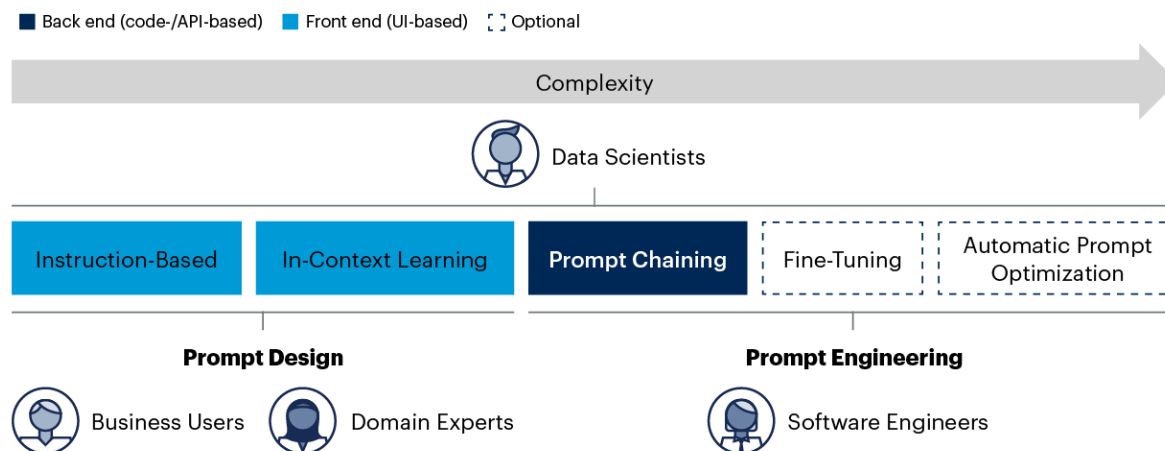
Generative AI models, and in particular LLMs, are pretrained, so data that enterprises want to use with these models cannot be added to the training set. Instead, prompts can be used to feed content to the model with an instruction to carry out a function. They can also be fine-tuned by providing examples of specific tasks.

The ever-increasing number of generative AI models means that data scientists are best-placed to understand the capabilities and limits of models, and to determine whether to pursue a purely prompt-based or fine-tuning based approach (or combination of approaches) for customization. The ultimate goal is to use machine learning itself to generate the best prompts and achieve automated prompt optimization.¹ This is in contrast to an end user of an LLM who concentrates on prompt design to manually alter prompts to give better responses.

The range of approaches for prompt-based tasks and roles involved are shown in Figure 1.

Figure 1: Breakdown of Prompt Engineering and Roles Involved

Breakdown of Prompt Engineering and Roles Involved



Source: Gartner
794952_C

Gartner

Prompt-based approaches involve both prompt design to provide clear definitions of what is required and back-end prompt engineering to automate, chain and configure prompts in a pipeline for use in AI applications. New styles of prompt techniques are being developed by data scientists in academia and industry. Examples include:

- Chain-of-Thought prompting, ² where a model is encouraged to show its working or thinking step-by-step; and
- Tree of Thoughts, where multiple prompts are evaluated and refined until a task is complete. ³

Data scientists can use prompt design and engineering to unlock capabilities of an LLM that are not well-known or even thought to be possible.

Data scientists will also need to decide when an LLM can be used to go beyond text generation to the execution of processes and utilization of external services. In April 2023, Gartner's social media analysis found a significant number of conversations from business executives related to AutoGPT, an open-source application that automates tasks by leveraging GPT4 and integrations to tools. Utilizing models in this way will require significant prompt engineering.

The APIs that interact with commercially available models have a number of parameters that can be adjusted (such as temperature, bias and penalties) that can alter outputs depending on the use case. Data scientists can reason on the effect of parameter changes against different types of models.

Leveraging Prompt Engineering for Data Science and Machine Learning Tasks

Foundation models, and LLMs in particular, have become popular for creative tasks that require text or image content to be generated. LLMs, with their capability for natural language processing and computer vision, could also be used for named entity recognition or sentiment analysis, either through prompt-based or fine-tuning approaches. This makes the data scientist's task of building a machine learning model for predictions much easier than standard approaches.

Prompt engineering can help develop systems that make use of data on which the model has not been trained. A current approach known as retrieval augmented generation (RAG) combines search and retrieval systems, such as insight engines and vector databases, with LLMs to answer questions and produce outputs. Implementing this approach requires knowledge of both forming correct prompts and utilizing other capabilities of the model, such as generating embeddings.

Code Generation With Prompt Engineering

Some generative AI models have been trained specifically on code, which makes them usable for code generation, testing, documentation and bug fixing. Data scientists will use prompt engineering to accelerate the model building process, especially when using open-source packages and generating boilerplate code. Generative AI models can also assist in the initial phases of business understanding to validate measurement KPIs, feature selection and test strategies. All of these tasks will become common in developer tools that data scientists use.

The advent of models that can execute code, and not just generate it, also brings the ability to design more complex patterns that involve long-term and contextual memory, reasoning and retrieval augmented generation. The ease of creating these systems in a declarative way lowers the barrier for citizen data scientists or even those from nontechnical backgrounds to participate in delivering these systems, but data scientists possess the required knowledge on the fundamental techniques to be their lead composers and orchestrators.

Responsibility for Prompt Validation and Testing

Data scientists have been grounded in testing and validating predictive models for accuracy using a plethora of statistical scores and metrics. Prompt validation and testing will need to utilize adaptations of these measures to extensively test generative AI systems. With the increasing choice of generative AI models available. Prompts will also need to be tested against different models to find which ones perform better on specific tasks.

Finally, data scientists will need to ensure that prompt testing is protecting data privacy and that prompts have been optimally designed for the model being used. This closely resembles the testing for bias and unintended discrimination that is now a standard part of responsible AI practices. Data scientists will need to extend and expand responsible AI practices for generative AI models and be a part of education and AI literacy programs for knowledge workers and end users who use the models as part of their work.

Recommended by the Authors

[Quick Answer: How Can You Manage Trust, Risk and Security for ChatGPT Usage in Your Enterprise?](#)

[Innovation Insight for ML-Powered Coding Assistants](#)

[AI Design Patterns for Large Language Models](#)

[What Are the Essential Roles for Data and Analytics?](#)

Evidence

Approved Methodology: Gartner conducts social listening analysis leveraging third-party data tools to complement or supplement the other fact bases presented in this document. Due to its qualitative and organic nature, the results should not be used separately from the rest of this research. No conclusions should be drawn from this data alone. Social media data in reference is from 1 November 2022 to 30 April 2023, in all geographies (except China) and recognized languages.

¹ [Automatic Prompt Optimization with “Gradient Descent” and Beam Search](#), Microsoft

² [\[2201.11903\] Chain-of-Thought Prompting Elicits Reasoning in Large Language Models](#)

³ [\[2305.10601\] Tree of Thoughts: Deliberate Problem Solving With Large Language Models](#)

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by [Gartner's Usage Policy](#). Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "[Guiding Principles on Independence and Objectivity](#)." Gartner research may not be used as input into or for the training or development of generative artificial intelligence, machine learning, algorithms, software, or related technologies.