Intents Are Back!
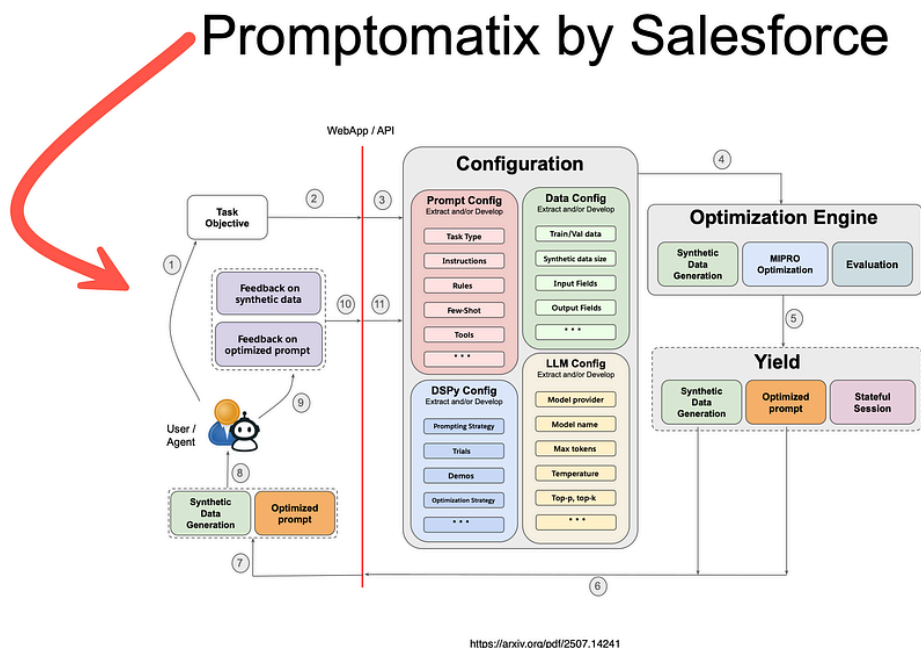
There has been a flurry of activity around prompt optimisation, from OpenAI and also now Salesforce…

Author: Cobus Greyling | 6 min read | Jul 30, 2025
URL: https://cobusgreyling.medium.com/intents-are-back-f361c7141653

There is a renewed focus on optimising prompts…as I have mentioned before, OpenAI is focussing on prompt optimisation for deep research inference…

…and also matching use-cases with the most appropriate model.

There has been research in the past that focussed on the idea of an interface where users do not perform prompting, but rather express their intent and expectations and the prompt is written and optimised for them.

Promptomatix

Recent research from Salesforce introduced Promptomatix…which can be described as an ***automatic prompt optimisation framework*** that transforms ***natural language task descriptions*** into ***high-quality prompts*** without requiring manual tuning or domain expertise.

The system analyses:

- User intent,
- generates synthetic training data, s
- elects prompting strategies, and
- refines prompts using cost-aware objectives.

System Analysis & Prompt Refinement Process...

Analyze User Intent

Generate Synthetic Data

Select Prompting Strategies

Refine Prompts

Intents

I love the fact that intents remain important….

Intents are really the goal or the purpose of the user, capturing this accurately removes ambiguity and leads to an accurate response.

In the chatbot world of old, NLU models consisted of intents (verbs) and entities (nouns). And intents were really a label for different classifications of pre-defined user queries.

Now in the Generative AI world intents remain important as they evolve from *static* labels into *dynamic*, context-aware mechanisms that help Large Language Models (LLMs) to interpret user queries with greater nuance and flexibility.

Rather than relying solely on predefined classifications, generative AI leverages advanced intent detection to infer real-time motivations.

Capturing **user intent** with the help of Generative AI involves the interpretation and classify the underlying goals behind user queries or conversations, transforming ambiguous inputs into precise, actionable insights.

This process has evolved from traditional chatbot frameworks, where intents were predefined labels for query classification, to dynamic systems powered by large language models (LLMs), which analyse contextual embeddings and more to predict motivations in real-time.

Disambiguation plays a critical role here, addressing linguistic ambiguities.
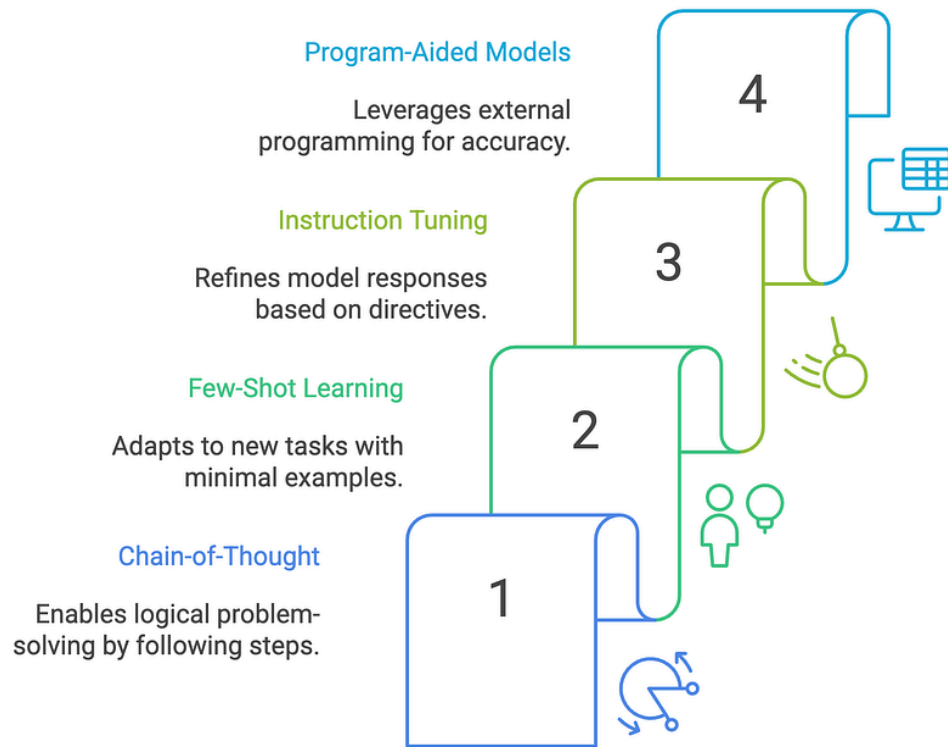
Prompt Engineering

The field has witnessed the rapid development of sophisticated prompting techniques, including…

- **Chain-of-Thought** reasoning,
- Few-shot learning,
- Instruction tuning, and
- Program-aided language models.

However, scalable prompt engineering remains notably absent, hindering broader accessibility and practical deployment.

Prompt Engineering

Program-Aided Models

Leverages external programming for accuracy.

4

Instruction Tuning

Refines model responses based on directives.

3

Few-Shot Learning

Adapts to new tasks with minimal examples.

2

Chain-of-Thought

Enables logical problem-solving by following steps.

1

Current Limitations

According to Salesforce, current **prompt engineering** practices face several fundamental challenges that limit their scalability, accessibility and practical deployment.

First, crafting effective prompts demands specialised knowledge of LLM behaviour, advanced prompting techniques (Tree-of-Thought, Program-of-Thought & ReAct) and domain-specific optimisation strategies.

This creates a significant expertise barrier for domain experts lacking technical machine learning knowledge, thereby hindering the democratisation of LLM capabilities across diverse user communities.
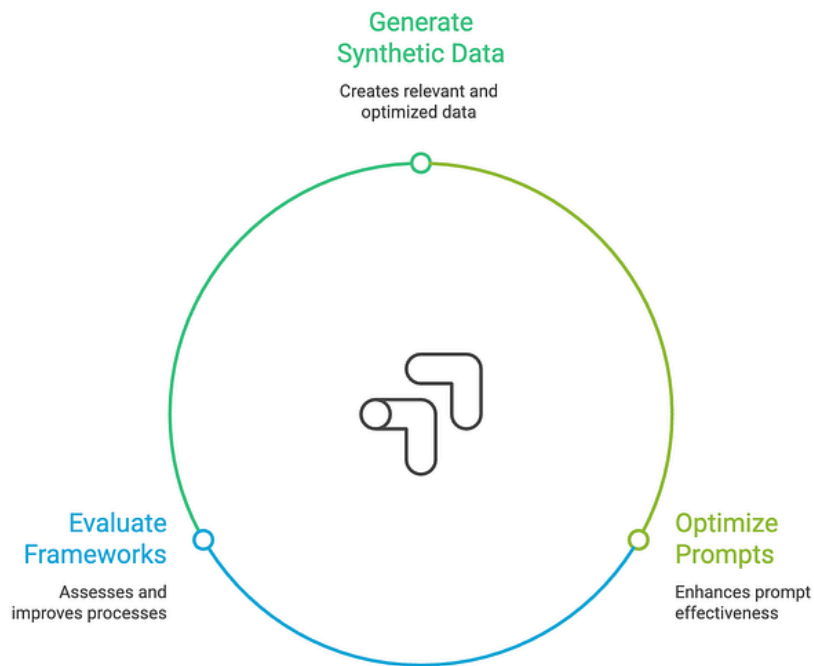
Second, LLMs exhibit high *sensitivity* to prompt variations, resulting in unpredictable outputs that can fluctuate dramatically due to minor changes in wording, formatting, or example selection.

This instability complicates the development of robust, production-ready applications that demand consistent performance across varied inputs and contexts.

Third, inefficient prompts consume excessive computational resources, leading to elevated costs and latency (especially for research tasks))without commensurate performance gains.

Manual optimisation often overlooks systematic cost-performance trade-offs, resulting in suboptimal resource utilisation in large-scale deployments.

Optimisation Engine Cycle

Generate Synthetic Data — Creates relevant and optimized data

Optimize Prompts — Enhances prompt effectiveness

Evaluate Frameworks — Assesses and improves processes

The Goal

The goal of the researchers was to present a zero-configuration framework that automates the full prompt optimisation pipeline — from *intent* analysis to performance evaluation — using only natural language task descriptions.
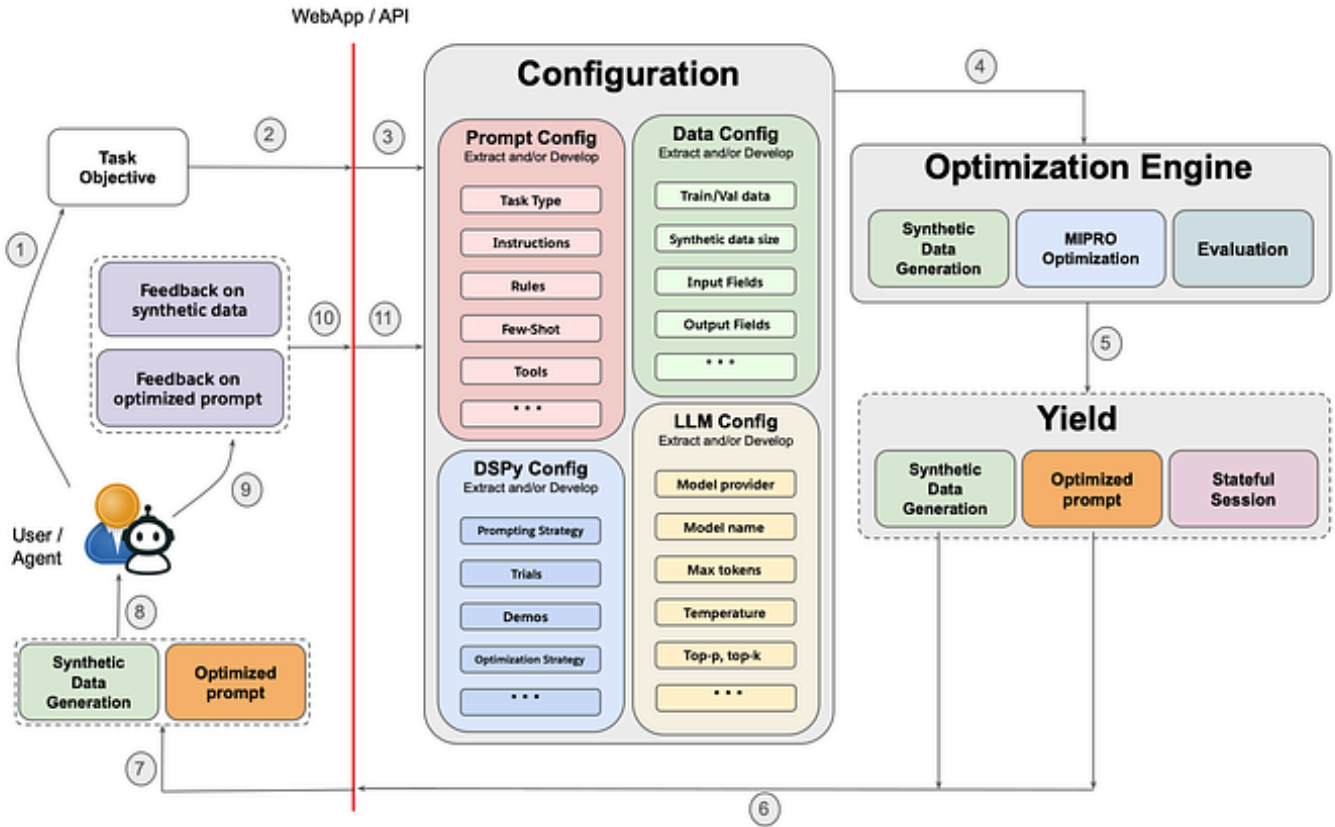
They introduce novel techniques for intelligent synthetic data generation that eliminate data bottlenecks in prompt optimisation.

Salesforce propose a cost-aware optimisation objective that balances quality with computational efficiency, enabling user-controlled trade-offs.

Limitations of Current Approaches

The analysis reveals common limitations across existing frameworks:

1. Manual configuration requirements for technique selection and parameter tuning, creating barriers for users without deep technical expertise in prompt engineering methodologies,
2. Lack of synthetic data generation capabilities, forcing users to manually collect and curate task-specific training datasets which is time-consuming and resource-intensive.
3. Limited end-to-end automation, requiring fragmented workflows with substantial manual coordination between different optimisation stages.
4. Technical complexity barriers for non-expert users, as most tools require programming knowledge and understanding of underlying optimisation algorithms.
5. Absence of cost-aware optimisation strategies that systematically balance performance improvements with computational efficiency and resource costs.
6. Lack of unified interfaces across different optimisation backends, creating vendor lock-in and limiting flexibility in choosing appropriate strategies for specific tasks.
7. Insufficient user feedback integration mechanisms, preventing iterative refinement based on domain-specific requirements and real-world deployment experiences.

Synthetic Data

The Synthetic Data Generation module tackles one of the primary bottlenecks in prompt optimisation by automatically producing high-quality, task-specific training datasets.

It implements a sophisticated multi-stage process, starting with template extraction from sample data, followed by intelligent batch generation that adheres to token limits and promotes diversity.

Table 3: Feature Comparison with Existing Frameworks

Framework	Auto Data	Auto Technique	Auto Metric	Zero Config	Feedback	Cost Opt	Prompt Mgmt
DSPy	✗	✗	✗	✗	✗	✓	✗
AdalFlow	✗	✗	✗	✗	✗	✓	✓
Promptify	✗	✗	✗	✗	✗	✗	✗
LangChain Prompt Canvas	✗	✗	✗	✗	✓	✓	✓
PromptWizard	✓	~	✗	✗	✗	✓	✗
Promptomatix	✓	✓	✓	✓	✓	✓	✓

The generation process leverages advanced prompting techniques to produce examples encompassing various complexity levels, edge cases, and stylistic variations, while preserving consistency with task requirements.

This method overcomes traditional data collection challenges and supports optimisation in specialised domains where training data is limited.

Mentions by Author

- [arxiv.org - Promptomatix: An Automatic Prompt Optimization Framework for Large Language Models - Large Language Models (LLMs) perform best with well-crafted prompts, yet prompt engineering remains manual…](#)
- [github.com - GitHub - SalesforceAIResearch/promptomatix: An Automatic Prompt Optimization Framework for Large… - An Automatic Prompt Optimization Framework for Large Language Models - SalesforceAIResearch/promptomatix](#)
- [Using AI Agents For Prompt Optimisation & Language Model Selection - Orchestrating AI Agents for prompt improvement & matching the best model to the use-case…](#)
- [cookbook.openai.com - Optimize Prompts | OpenAI Cookbook - Crafting effective prompts is a critical skill when working with AI models. Even experienced users can inadvertently…](#)
- [OpenAI Multi-AI Agent Research Framework - If you want to gain a better understanding of how a multi-AI Agent system look in practical terms, this will help you.](#)
- [OpenAI Deep Research AI Agent Architecture - Recently OpenAI shows their ideal scenario for creating a Deep Research AI Agent…](#)
- [github.com - openai-cookbook/examples/Optimize_Prompts.ipynb at main · openai/openai-cookbook - Examples and guides for using the OpenAI API. Contribute to openai/openai-cookbook development by creating an account…](#)