

DSPy - Programming—not prompting—LMs

DSPy is the framework for programming—rather than prompting—language models. It allows you to iterate fast on building modular AI systems and offers algorithms for optimizing their prompts and weights, whether you're building simple classifiers, sophisticated RAG pipelines, or Agent loops.

DSPy stands for Declarative Self-improving Python. Instead of brittle prompts, you write compositional Python code and use DSPy to teach your LM to deliver high-quality outputs.

Getting Started I: Install DSPy and set up your LM

```
> pip install -U dspy
```

Local LMs on your laptop

First, install Ollama and launch its server with your LM.

```
> curl -fsSL https://ollama.ai/install.sh | sh
```

```
> ollama run llama3.2:1b
```

Then, connect to it from your DSPy code.

```
import dspy
```

```
lm = dspy.LM('ollama_chat/llama3.2', api_base='http://localhost:11434', api_key='')
```

```
dspy.configure(lm=lm)
```