The paradigm shift from MLOps to LLMOps

	Traditional MLOps	LLMOps
Target audiences	ML Engineers Data Scientists	ML Engineers App developers
Assets to share	Model, data, environments, features	LLM, agents, plugins, prompts, APIs
Metrics/evaluations	Accuracy	**Quality:** accuracy, similarity **Harm:** bias, toxicity **Correct:** groundness **Cost:** token per request **Latency:** response time, RPS
ML models	Build from scratch	Pre-built, fine-tuned served as API (MaaS)

Implementing LLMOps in Azure AI & Azure AI Studio

Discover foundation models with ease through a unified model catalog

Access and customize thousands of foundation models

Scale LLM practices with precision to ensure consistent and streamlined workflows

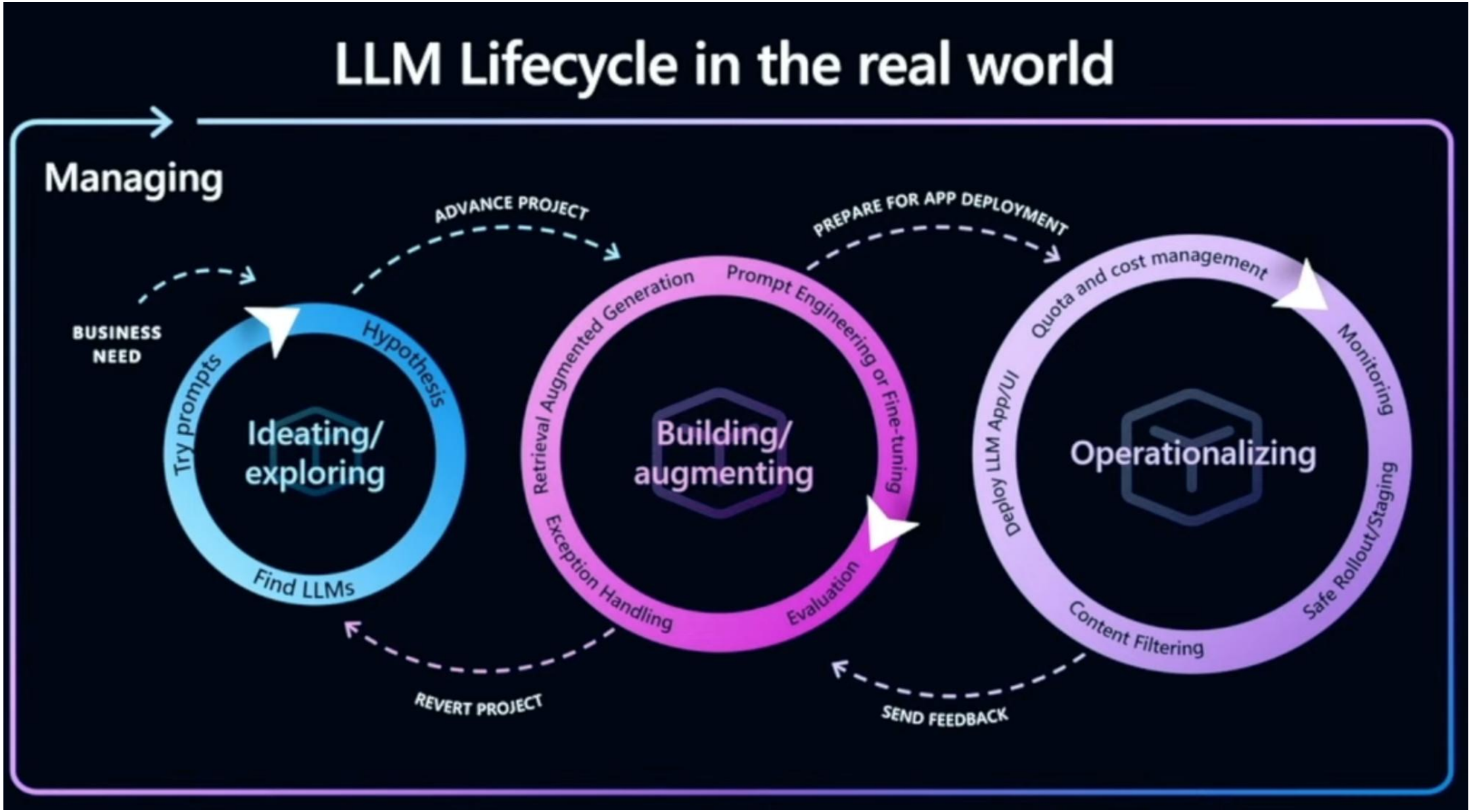
Version-control prompts and integrate with CI/CD

Foster team collaboration among various types of roles – data engineers, ML engineers, app developers

Promote, discover, and share LLM assets across organizations

Cultivate trust and transparency with built-in AI governance and Responsible AI

Engineered to be secure, trusted and responsible


LLM Lifecycle in the real world
Managing
BUSINESS NEED
ADVANCE PROJECT
PREPARE FOR APP DEPLOYMENT
Ideating/ exploring
Try prompts
Hypothesis
Find LLMs
Building/ augmenting
Retrieval Augmented Generation
Prompt Engineering or Fine-tuning
Exception Handling
Evaluation
Operationalizing
Quota and cost management
Monitoring
Safe Rollout/Staging
Content Filtering
Deploy LLM App/UI
REVERT PROJECT
SEND FEEDBACK

Grounding your LLM apps with your data

Fine-tuning

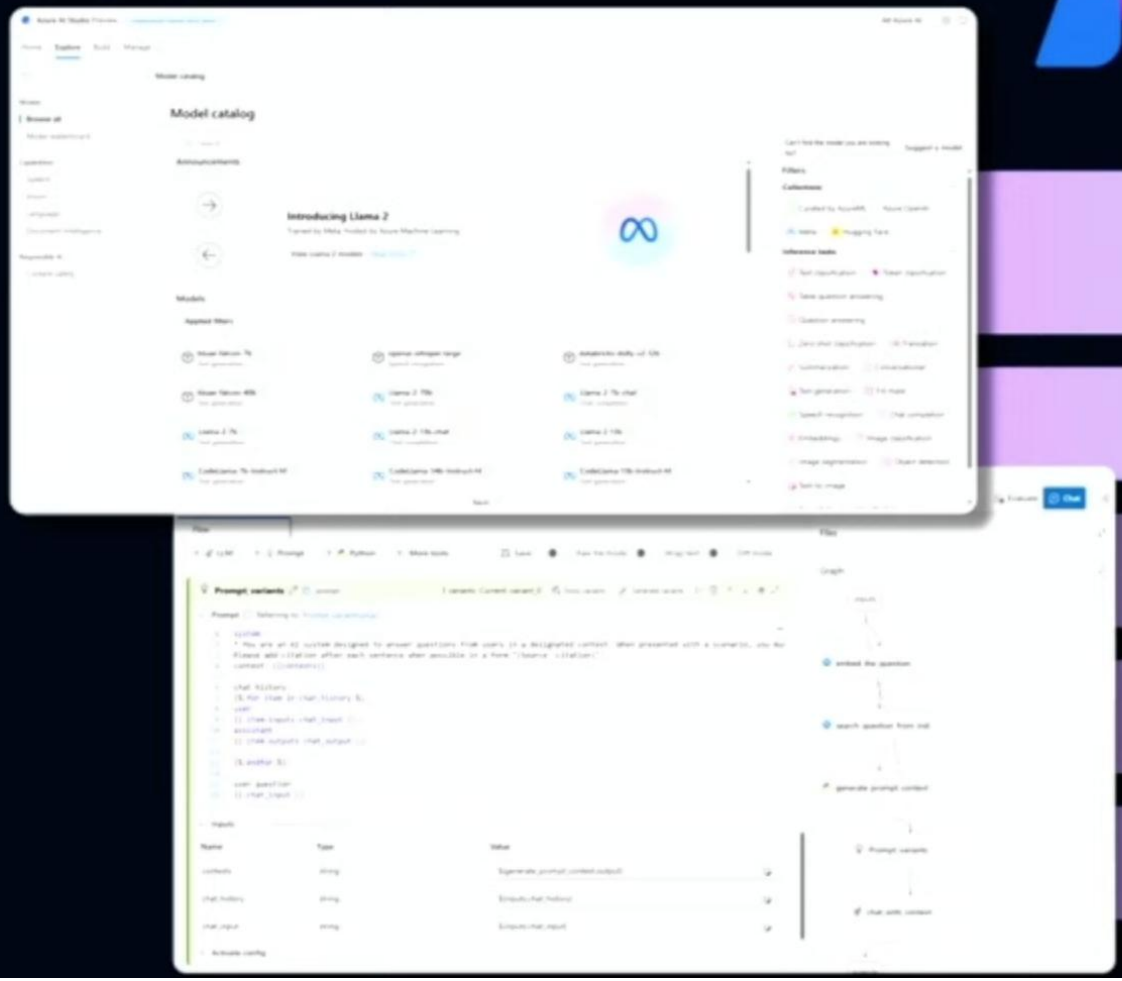
-> complex, expensive

RAG - Retrieval Augmented Generation

-> enriching a prompt with context

-> less complex than fine tuning a model

Azure AI Studio is a platform for Generative AI



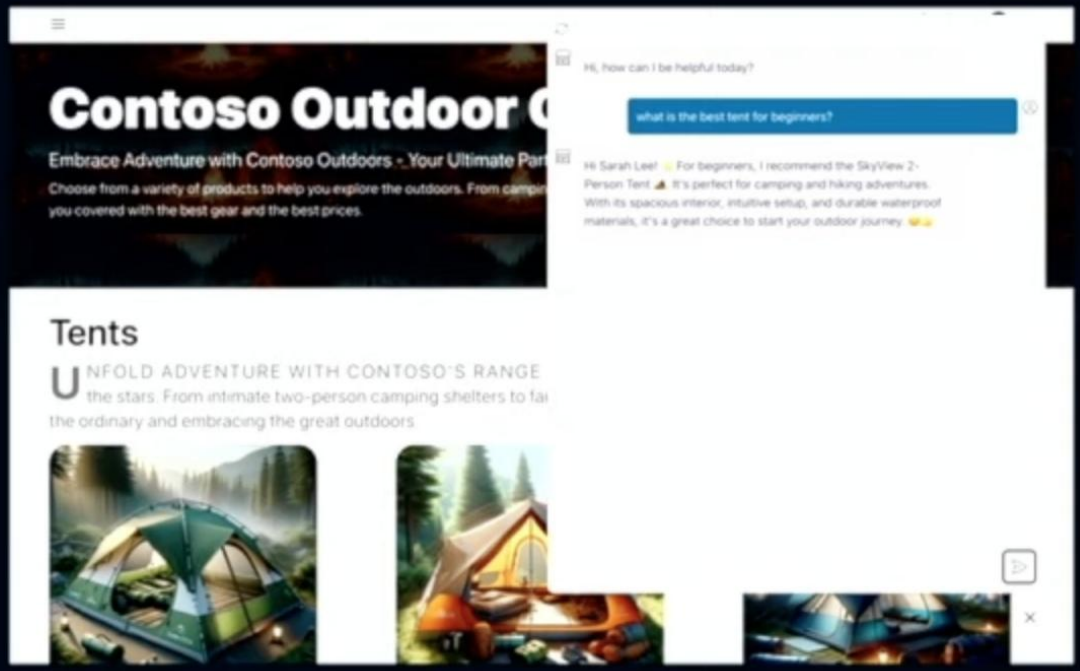
- Access to thousands of LLMs from OpenAI, Meta, Hugging Face
- Data grounding with RAG
- Prompt engineering/evaluation
- Built-in safety and responsible AI
- Continuous monitoring for LLMs

Use case - Contoso customer service chat

Use case risk

- High risk = Collaborate
  - Patient diagnostics
  - Economic decisions

- Low risk = Automate
  - Customer support (with handover)
  - Information retrieval
  - Content generation
  - Language translation
  - Education



Azure AI prompt flow

Connections

Manage APIs and external data sources

- Seamless integration with pre-built LLMs like Azure OpenAI Service
- Built-in safety system with Azure AI Content Safety
- Effectively manage credentials or secrets for APIs
- Create your own connections in Python tools

Orchestrate flows

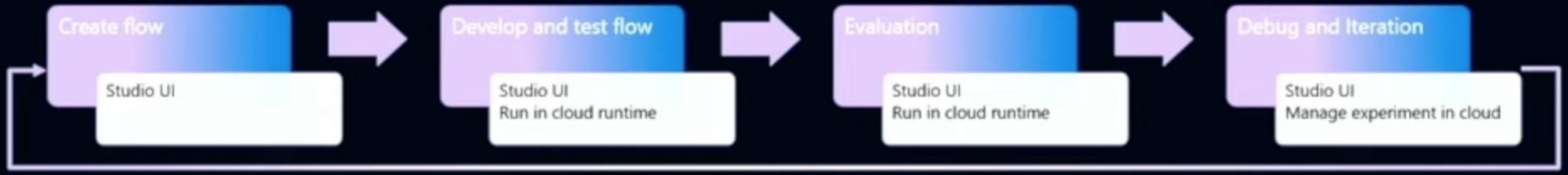
Develop your LLM flow from scratch

- Use any framework such as LangChain or Semantic Kernel to build initial flows
- Add your own reusable tools
- Manage your flows as files
- Track run history

Prompt flow transition journey (cloud - local)

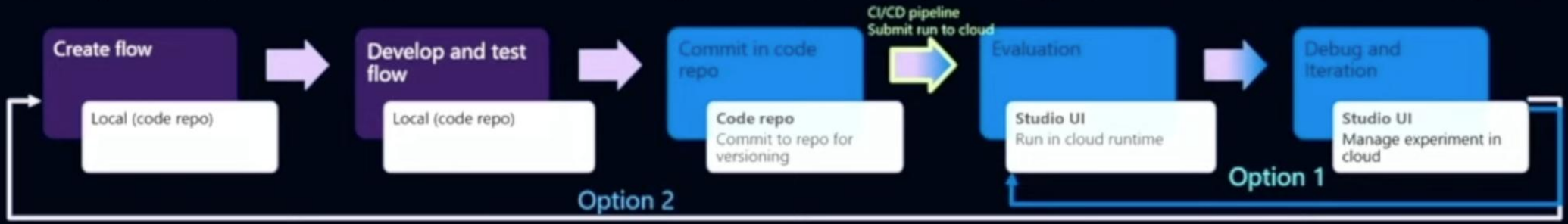
Journey 1: Pure Cloud UI Cloud develop, Cloud run, Cloud manage

Non-developers who prefer to develop flow and experimentally run flow with pure UI experience in workspace.



Journey 2: Local + cloud Cloud/repo develop, Repo versioning, Repo submit run to cloud, Cloud manage

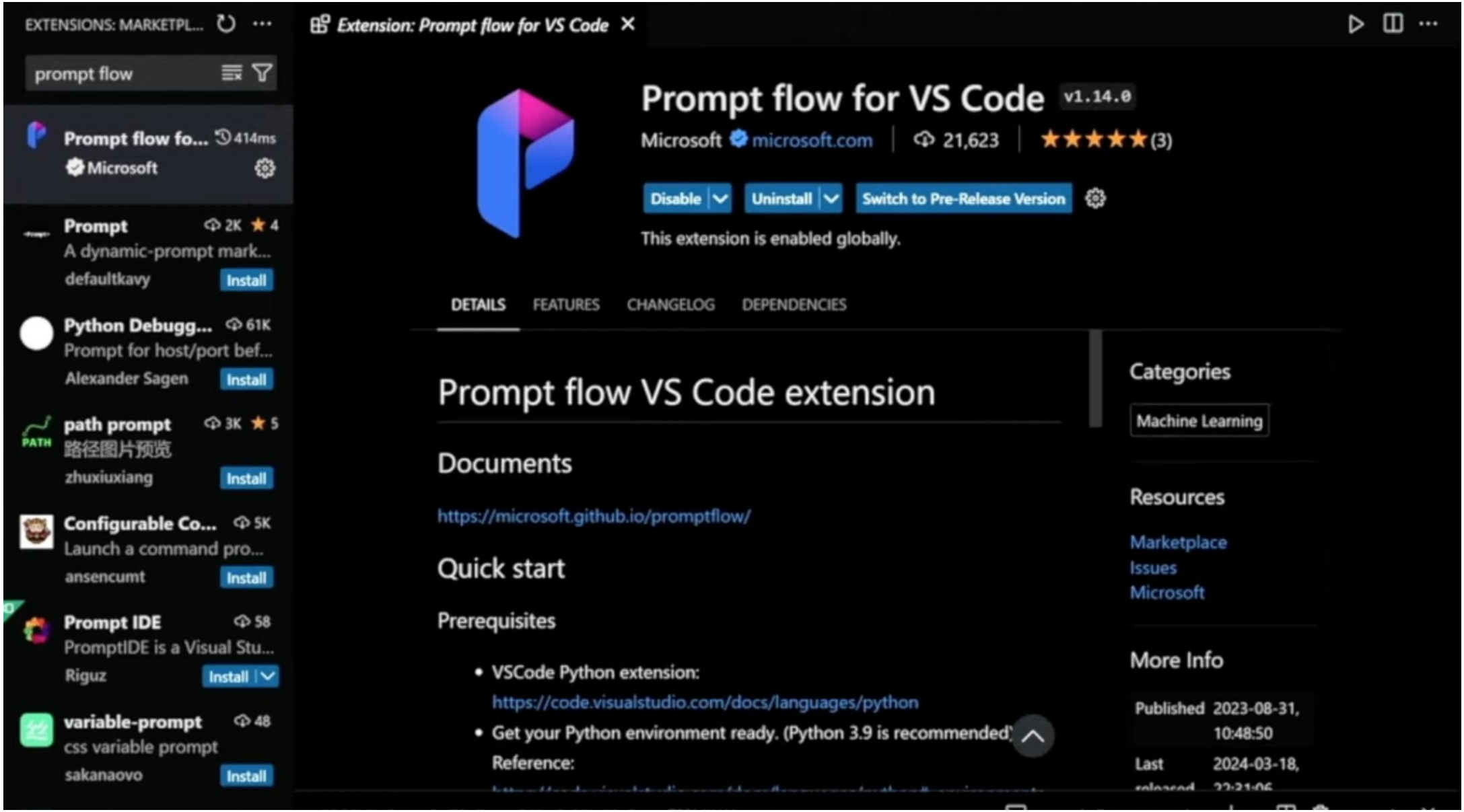
Developers who prefer to develop flow in UI, then versioning in code repo, and experimentally run flow in code repo with CI/CD integration.

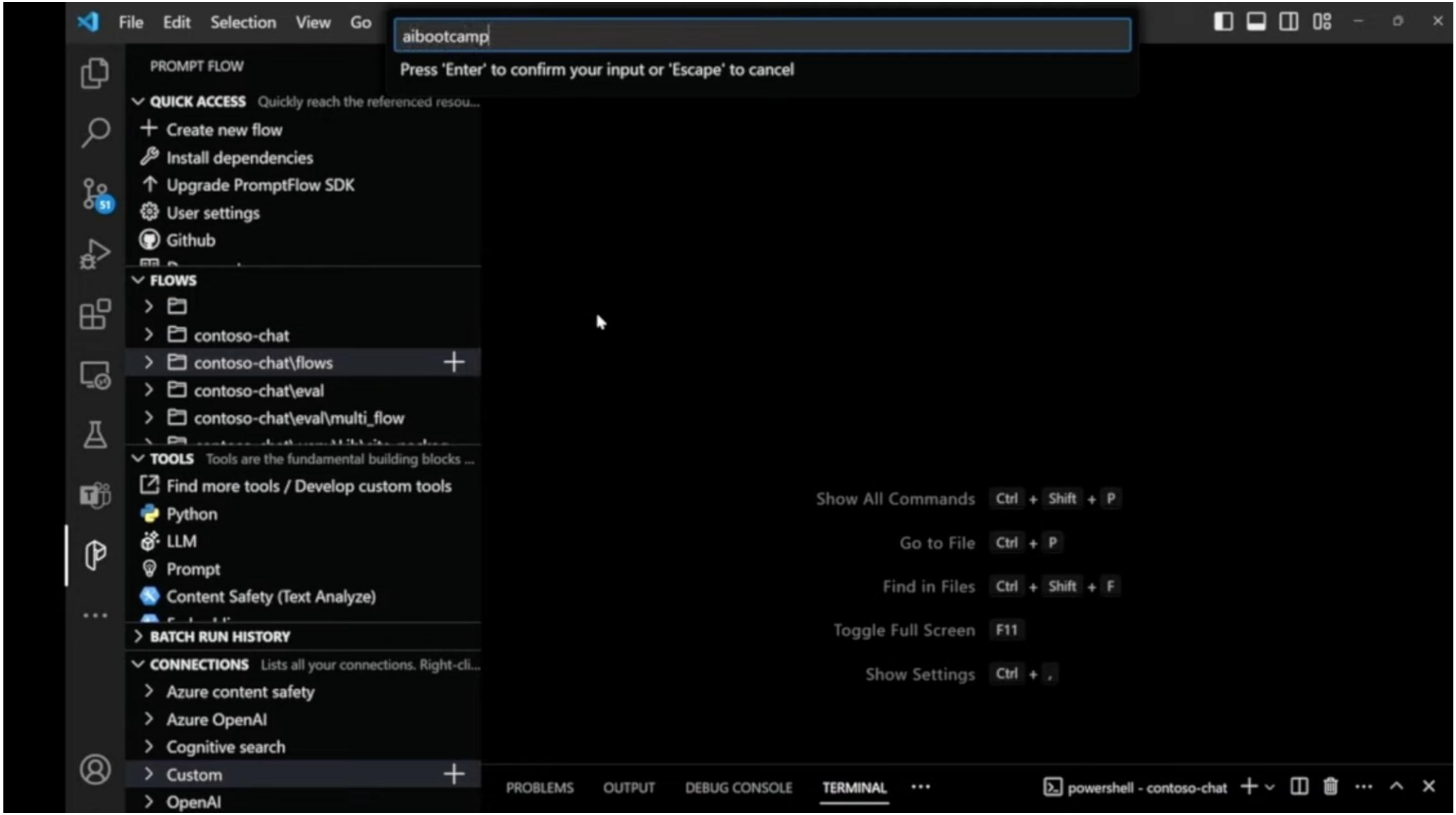


Tips for iteration:

Option 1: you can stay in cloud UI, edit flow based on experiment snapshot. Then download flow files to local to commit to repo.

Option 2: you can return to code repo, leverage VS code extension to edit flow locally. Then go through the subsequent processes.


EXTENSIONS: MARKETPL...
prompt flow
Prompt flow fo... 414ms
Microsoft
Prompt 2K 4
A dynamic-prompt mark...
defaultkavy
Install
Python Debugg... 61K
Prompt for host/port bef...
Alexander Sagen
Install
path prompt 3K 5
路径图片预览
zhuxiuxiang
Install
Configurable Co... 5K
Launch a command pro...
ansencumt
Install
Prompt IDE 58
PromptIDE is a Visual Stu...
Riguz
Install
variable-prompt 48
css variable prompt
sakanaovo
Install
Extension: Prompt flow for VS Code
Prompt flow for VS Code v1.14.0
Microsoft microsoft.com
21,623
(3)
Disable
Uninstall
Switch to Pre-Release Version
This extension is enabled globally.
DETAILS
FEATURES
CHANGELOG
DEPENDENCIES
Prompt flow VS Code extension
Documents
https://microsoft.github.io/promptflow/
Quick start
Prerequisites
VSCode Python extension:
https://code.visualstudio.com/docs/languages/python
Get your Python environment ready. (Python 3.9 is recommended)
Reference:
Categories
Machine Learning
Resources
Marketplace
Issues
Microsoft
More Info
Published 2023-08-31, 10:48:50
Last released 2024-03-18, 22:21:06


File Edit Selection View Go
aibootcamp
Press 'Enter' to confirm your input or 'Escape' to cancel
PROMPT FLOW
QUICK ACCESS Quickly reach the referenced resou...
Create new flow
Install dependencies
Upgrade PromptFlow SDK
User settings
Github
FLOWS
contoso-chat
contoso-chat\flows
contoso-chat\eval
contoso-chat\eval\multi_flow
TOOLS Tools are the fundamental building blocks ...
Find more tools / Develop custom tools
Python
LLM
Prompt
Content Safety (Text Analyze)
BATCH RUN HISTORY
CONNECTIONS Lists all your connections. Right-cli...
Azure content safety
Azure OpenAI
Cognitive search
Custom
OpenAI
Show All Commands Ctrl + Shift + P
Go to File Ctrl + P
Find in Files Ctrl + Shift + F
Toggle Full Screen F11
Show Settings Ctrl + ,
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
powershell - contoso-chat