

Related Roadmaps

✓ AI Engineer Roadmap

✓ AI and Data Scientist Roadmap

✓ MLOps Roadmap

✓ AI Red Teaming Roadmap

✓ Prompt Engineering Roadmap

Find the detailed version of this roadmap along with other similar roadmaps

roadmap.sh

Streamed vs Unstreamed Responses

Reasoning vs Standard Models

Fine-tuning vs Prompt Engineering

Embeddings and Vector Search

Understand the Basics of RAG

Pricing of Common Models

AI Agents

Learn the Pre-requisites

Basic Backend Development

Git and Terminal Usage

REST API Knowledge

Backend Beginner Roadmap

Git and GitHub Roadmap

API Design Roadmap

Understand the Basics

LLM Fundamentals

Open Weight Models

Closed Weight Models

Transformer Models and LLMs

Model Mechanis

Tokenization

Context Windows

Token Based Pricing

Generation Controls

Temperature

Top-p

Frequency Penalty

Presence Penalty

Stopping Criteria

Max Length

Model Families and Licences

1 Perception / User Input

2 Reason and Plan

3 Acting / Tool Invocation

4 Observation & Reflection

AI Agents 101

What are AI Agents?

What are Tools?

Agent Loop

Example Usecases

Personal assistant

Code generation

Data analysis

Web Scraping / Crawling

NPC / Game AI

What is Prompt Engineering

Prompt Engineering

Writing Good Prompts

Be specific in what you want

Provide additional context

Use relevant technical terms

Use Examples in your Prompt

Iterate and Test your Prompts

Specify Length, format etc

Prompt Engineering Roadmap

Tools / Actions

Tool Definition

Name and Description

Input / Output Schema

Error Handling

Usage Examples

Examples of Tools

Web Search

Code Execution / REPL

Database Queries

API Requests

Email / Slack / SMS

File System Access

Model Context Protocol (MCP)

Core Components

MCP Hosts

MCP Client

MCP Servers

Creating MCP Servers

Deployment Modes

Local Desktop

Remote / Cloud

Agent Memory

What is Agent Memory?

Maintaining Memory

RAG and Vector Databases

User Profile Storage

Summarization / Compression

Forgetting / Aging Strategies

Episodic vs Semantic Memory

Short Term Memory

Within Prompt

Long Term Memory

Vector DB / SQL / Custom

Agent Architectures

Evaluation and Testing

Metrics to Track

Unit Testing for Individual Tools

Integration Testing for Flows

Human in the Loop Evaluation

Frameworks

LangSmith

DeepEval

Ragas

Debugging and Monitoring

Structured logging & tracing

Observability Tools

LangSmith

Helicone

LangFuse

openllmetry

Common Architectures

RAG Agent

ReAct (Reason + Act)

Chain of Thought (CoT)

Other Architecture Patterns

Planner Executor

DAG Agents

Tree-of-Thought

Building Agents

Manual (from scratch)

Direct LLM API calls

Implementing the agent loop

Parsing model output

Error & Rate-limit handling

Building Using Frameworks

Langchain

LlamaIndex

Haystack

AutoGen

CrewAI

Smol Depot

LLM Native "Function Calling"

OpenAI Functions Calling

OpenAI Assistant API

Gemini Function Calling

Anthropic Tool Use

Security & Ethics

Prompt Injection / Jailbreaks

Tool sandboxing / Permissioning

Data Privacy + PII Redaction

Bias & Toxicity Guardrails

Safety + Red Team Testing

Visit the following relevant tracks

AI Engineer

AI & Data Scientist