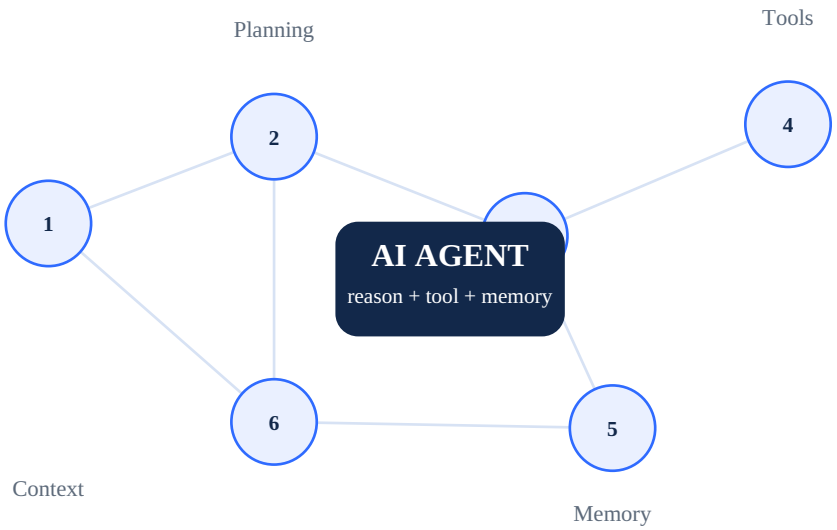
AI AGENTS

Từ Large Language Models đến những hệ thống có khả năng lập kế hoạch, sử dụng công cụ, ghi nhớ và tự thực thi



Bản tổng hợp tiếng Việt dựa trên 8 trang trong Prompt Engineering Guide của DAIR.AI

Phiên bản biên soạn: 01/10/2026

LỜI MỞ ĐẦU

Tại sao cần một cuốn sách về AI Agents?

AI đang chuyển từ mô hình "hỏi - đáp" sang hệ thống có thể tự chia nhỏ mục tiêu, lựa chọn công cụ, thực hiện nhiều bước và kiểm tra kết quả. Các trang nguồn được cung cấp tập trung đúng vào sự chuyển dịch này. Cuốn sách này tái cấu trúc toàn bộ nội dung thành một mạch học tập liền mạch, giữ thuật ngữ chuyên ngành bằng tiếng Anh và diễn giải lại bằng tiếng Việt để có thể đọc nhanh, hiểu nhanh và áp dụng vào thiết kế hệ thống.

Cách đọc cuốn sách

Hãy xem AI Agent như một hệ thống chứ không chỉ là một prompt. Trước hết hiểu Agent là gì, sau đó hiểu các thành phần, phân biệt Workflow với Agent, học cách thiết kế Context, hiểu Function Calling và cuối cùng đi tới Deep Agents.

Lưu ý về nội dung và bản quyền

Đây là bản tổng hợp, diễn giải và cấu trúc hóa bằng tiếng Việt từ các trang công khai được cung cấp. Nội dung không sao chép nguyên văn toàn bộ nguồn. Các ví dụ trong sách được viết lại hoặc mở rộng để phục vụ mục đích học tập. Nguồn gốc được liệt kê ở cuối sách.

MỤC LỤC

Chương 1 - Hiểu AI Agent trong 10 phút

Định nghĩa, khác biệt với LLM thông thường, vì sao Agent hữu ích và các nhóm bài toán phù hợp.

Chương 2 - Ba thành phần cốt lõi của Agent

Planning, Tool Utilization và Memory.

Chương 3 - AI Workflows và AI Agents

Prompt Chaining, Routing, Parallelization, Agent Loop, lựa chọn Workflow/Agent và mô hình Hybrid.

Chương 4 - Context Engineering

Vì sao context quyết định hành vi, các lớp context, observability, validation và những lỗi thường gặp.

Chương 5 - Context Engineering Deep Dive

Case study Deep Research Agent, kiến trúc nhiều agent, thiết kế system prompt, quản lý context và error handling.

Chương 6 - Function Calling và Agent Loop

Cách LLM kết nối công cụ, schema, tham số, quan sát, debugging và best practices.

Chương 7 - Deep Agents

Planning dài hạn, orchestrator - sub-agent, external memory, context engineering và verification.

Chương 8 - Ghép tất cả thành một hệ thống

Một blueprint thực tế và checklist thiết kế production-ready.

Phụ lục - Thuật ngữ cốt lõi và nguồn tham khảo

CHƯƠNG 1

Hiểu AI Agent trong 10 phút

Agent là gì, tại sao xuất hiện và khi nào nên dùng

Từ câu hỏi sang hành động



Một AI Agent có thể được hiểu đơn giản là một hệ thống dùng LLM để nhận mục tiêu, suy luận về cách thực hiện và sử dụng các khả năng bổ sung để hoàn thành công việc. Điểm khác biệt quan trọng là Agent không dừng ở việc tạo văn bản: nó có thể hành động, quan sát kết quả và điều chỉnh bước tiếp theo.

Định nghĩa dễ nhớ

LLM trả lời câu hỏi. Agent theo đuổi một mục tiêu.

Agent khác LLM thông thường ở đâu?

- LLM thông thường thường xử lý một yêu cầu và tạo ra một đầu ra. Agent có thể thực hiện chuỗi bước nối tiếp nhau.

- Agent có thể truy cập dữ liệu hoặc hệ thống bên ngoài thay vì chỉ dựa vào kiến thức sẵn có trong model.

- Agent có thể sử dụng memory để duy trì trạng thái và nối tiếp công việc.

- Agent có khả năng planning và reflection, tức chia bài toán thành bước và xem lại tiến trình.

Vì sao Agent hữu ích?

Các tác vụ hẹp như dịch, viết email hoặc tóm tắt thường phù hợp với một lần gọi LLM. Nhưng những bài toán như nghiên cứu thị trường, lập kế hoạch sự kiện, chăm sóc khách hàng hoặc phân tích dữ liệu cần nhiều nguồn thông tin và nhiều bước. Agent lấp khoảng trống đó bằng cách kết hợp LLM với tools, memory và khả năng điều phối.

Ví dụ cụ thể: nghiên cứu đối thủ

Yêu cầu: "Hãy phân tích 5 đối thủ của công ty và đề xuất 3 cơ hội mới."

- Bước 1: Agent hiểu mục tiêu và tách thành các câu hỏi nghiên cứu.

- Bước 2: Agent tìm dữ liệu web hoặc cơ sở dữ liệu nội bộ.

- Bước 3: Agent lưu và tổng hợp phát hiện quan trọng.

- Bước 4: Agent kiểm tra chéo và xác định điểm thiếu.

- Bước 5: Agent viết báo cáo cuối cùng.

Các nhóm ứng dụng phổ biến

Nhóm	Ví dụ
Research	Tìm hiểu sâu một chủ đề, tổng hợp nguồn
Customer Support	Phân loại, tra cứu, xử lý yêu cầu
Reporting	Phân tích dữ liệu và tạo báo cáo
E-commerce / Booking	Tìm sản phẩm, quản lý đơn hoặc đặt lịch
Coding / Data	Viết, sửa, chạy code hoặc phân tích dữ liệu

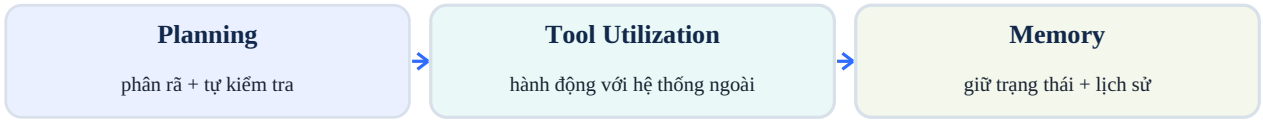
Nguồn chính: Prompt Engineering Guide: Agents / Introduction to AI Agents

CHƯƠNG 2

Ba thành phần cốt lõi của Agent

Planning - Tools - Memory

Agent = 3 năng lực phối hợp



Một Agent hữu ích không chỉ cần model mạnh. Nó cần ba khả năng phối hợp: planning để biết phải làm gì, tool utilization để thực sự làm được việc bên ngoài model, và memory để không mất thông tin giữa nhiều bước.

1. Planning - "Bộ não" của Agent

Planning là khả năng chuyển mục tiêu lớn thành chuỗi hành động, đánh giá tiến độ và điều chỉnh hướng đi khi thông tin mới xuất hiện. Nguồn nhấn mạnh các năng lực như task decomposition, self-reflection, adaptive learning và critical analysis.

Ví dụ

Mục tiêu "tìm 10 khách hàng tiềm năng" không nên được thực hiện như một bước duy nhất. Agent có thể xác định tiêu chí khách hàng, tạo danh sách nguồn, tìm kiếm, lọc, kiểm tra trùng lặp rồi mới tổng hợp.

2. Tool Utilization - "Tay chân" của Agent

Tool cho Agent quyền truy cập vào thế giới bên ngoài. Các nhóm tool điển hình gồm code execution, web search, calculator, image generation, database, API và các phần mềm nghiệp vụ. Điểm quan trọng không chỉ là có tool mà là model phải hiểu **khi nào** và **cách nào** để dùng tool.

3. Memory - "Bộ nhớ"

Memory giúp Agent duy trì continuity giữa các bước hoặc các phiên làm việc. Có thể hình dung hai lớp:

- **Short-term / Working Memory:** vùng đệm của phiên hiện tại, phục vụ in-context learning và duy trì mạch tác vụ.

- **Long-term Memory:** dữ liệu được lưu ngoài context, chẳng hạn vector store, database hoặc file, để có thể truy hồi về sau.

Một nguyên tắc quan trọng

Memory không phải là "lưu tất cả". Memory tốt là lưu và truy hồi đúng thứ có giá trị cho quyết định tiếp theo. Khi memory tăng quá nhanh, context trở nên dài, tốn token và khó điều khiển.

Nguồn chính: Prompt Engineering Guide: Agent Components

CHƯƠNG 3

AI Workflows và AI Agents

Biết khi nào cần quy trình cố định và khi nào cần tự chủ

Khía cạnh	AI Workflow	AI Agent
Control Flow	Định sẵn	Động, do LLM điều khiển
Decision Making	Logic được viết sẵn	Reasoning của LLM
Tool Usage	Code điều phối	Agent tự chọn
Adaptability	Cố định hơn	Linh hoạt hơn
Use Case	Bài toán xác định rõ	Bài toán mở, nhiều biến số

AI Workflow là gì?

Workflow là hệ thống trong đó các bước và đường đi đã được định nghĩa trước. Vì vậy nó dễ dự đoán, dễ kiểm thử và dễ kiểm soát. Đây là lựa chọn phù hợp khi yêu cầu ổn định và độ tin cậy sản xuất là ưu tiên.

Ba mẫu Workflow quan trọng

Prompt Chaining: chia một bài toán thành nhiều lần gọi LLM nối tiếp. Ví dụ: tạo outline -> kiểm tra -> mở rộng -> chỉnh sửa.

Routing: phân loại yêu cầu rồi chuyển đến chuỗi xử lý chuyên biệt. Ví dụ: câu hỏi chung -> general chain; hoàn tiền -> refund chain; lỗi kỹ thuật -> support chain.

Parallelization: chạy đồng thời nhiều tác vụ độc lập để giảm latency và tăng throughput. Ví dụ: kiểm tra nội dung theo nhiều tiêu chí song song.

Một Workflow có thể kiểm soát bằng điểm quyết định



AI Agent là gì trong bối cảnh này?

Agent thay control flow cố định bằng khả năng ra quyết định động. Agent có thể tự chọn tool, tự quyết định thứ tự gọi tool và điều chỉnh kế hoạch theo quan sát. Chính yếu tố đó tạo ra autonomy.

Khi nào nên dùng Workflow?

- Yêu cầu rõ và ổn định.

- Predictability và consistency quan trọng.