

Relatório 1

Vídeo: Fundamentos de Agentes de IA (I)

Kaíque Medeiros Lima

Introdução

Esse relatório analisa dois vídeos sobre a evolução e aplicação de agentes de Inteligência Artificial. O primeiro, *What are AI agents?*, apresenta os fundamentos de sistemas compostos e a transição de modelos isolados para agentes inteligentes. O segundo, *Andrew Ng Explores The Rise Of AI Agents And Agentic Reasoning | BUILD 2024 Keynote*, contextualiza tendências, práticas, padrões de design e demonstrações de workflows agenticos.

Descrição da Atividade

What are AI agents?

O vídeo começa destacando a limitação dos modelos monolíticos, treinados apenas sobre dados estáticos e a vantagem de *Compound AI*, em que um modelo de linguagem gera consultas para um banco de dados específico (por exemplo, dias de férias), recupera informações e formata respostas corretas.

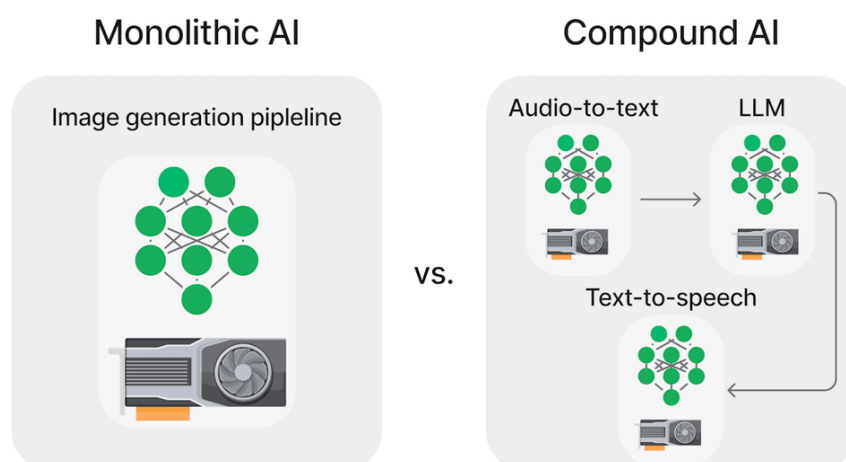


Figura 1: Diferença entre modelos monolíticos e sistemas compostos.

Em seguida, explica-se a arquitetura de um sistema composto: componentes modulares (modelos afinados, verificadores de saída, programas de decomposição de tarefas) e a facilidade de adaptação frente ao tuning pesado de modelos.

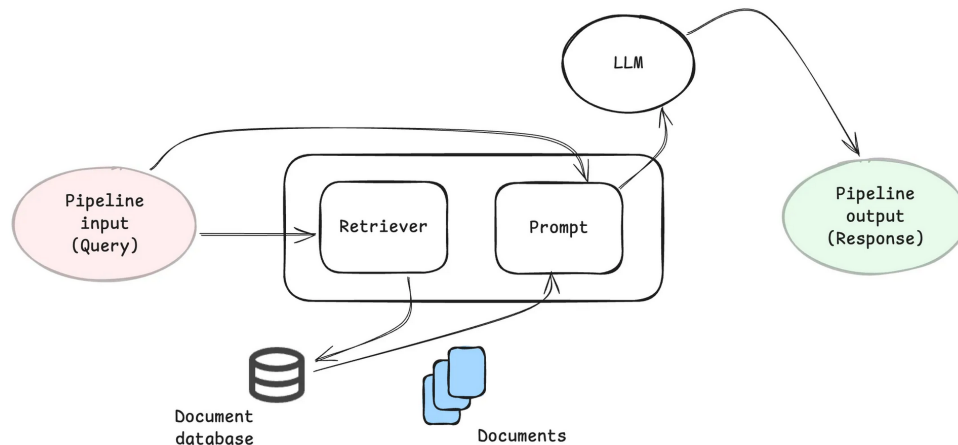


Figura 2: Arquitetura de um sistema composto.

Por fim, apresenta-se o conceito de agentes: grandes modelos comandando a lógica do sistema. Destacam-se três capacidades principais:

- **Raciocínio:** planejamento passo a passo (pensar devagar).
- **Ação:** invocação de ferramentas externas (buscas, calculadora, APIs).
- **Memória:** registro de logs internos e histórico de interações para personalização.

O framework ReACT (Reasoning and Acting) ilustra um agente que planeja, executa chamadas a ferramentas, observa resultados e itera até convergir em uma resposta confiável.

Andrew Ng Explores The Rise Of AI Agents And Agentic Reasoning

Andrew Ng inicia mapeando a *AI stack*: semicondutores, infraestrutura em nuvem (incluindo Snowflake), treinadores de modelos e, no topo, a camada de aplicações. Destaca o salto de prototipagem de meses para dias graças ao generative AI.

O foco recai em workflows agenticos como o próximo grande “motor” de inovação. Quatro padrões de design são detalhados:

1. **Reflection:** o modelo gera código, autocrítica e revisão iterativa.
2. **Tool Use:** chamadas de função e control APIs conduzidas pelo LLM.
3. **Planning:** decomposição de tarefas em sequência de ações (ex.: visão computacional passo a passo).
4. **Multi-Agent Collaboration:** agentes especializados colaboram para resolver sub-tarefas.

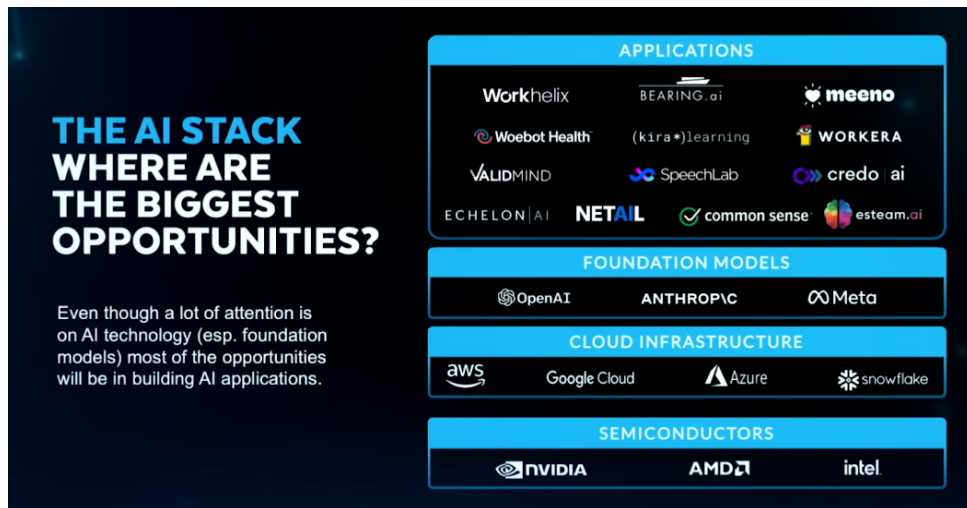


Figura 3: AI Stack

Por fim, Andrew Ng demonstra a *Vision Agent*, que conta jogadores em imagens e processa vídeos em cliques anotados via código gerado pelo agente, mostrando aplicações reais de IA multimodal.

Dificuldades

- Distinguir claramente *Compound AI* (sistemas compostos programáticos) de abordagens realmente agênticas.
- Compreender o fluxo de controle interno: como o modelo gera consultas, interpreta resultados e itera (ReACT).
- Aprofundar os quatro padrões de design agêntico e saber quando usar cada um.

Conclusão

Pelo o que entendi, o primeiro vídeo explica o que são agentes de IA e fundamenta as necessidades de sistemas além de modelos isolados, já no segundo vídeo, Andrew Ng expande esse conceito à prática, mostrando como workflows agênticos remodelam prototipagem e aplicações visuais.

Referências

- [1] “What are AI agents?”, 2024. Disponível em: <https://www.youtube.com/watch?v=F8NKVhkZZWI>
- [2] Ng, Andrew. “Andrew Ng Explores The Rise Of AI Agents And Agentic Reasoning | BUILD 2024 Keynote”, 2024. Disponível em: <https://www.youtube.com/watch?v=KrRD7r7y7NY>

- [3] Baseten. “Compound AI Systems Explained.” Baseten Blog, 2023. Disponível em: <https://www.baseten.co/blog/compound-ai-systems-explained/#what-is-a-compound-ai-system>
- [4] “What is Compound AI?”, 2024. Disponível em: <https://www.deepset.ai/blog/compound-ai>