

Relatório 1 - Vídeo: Fundamentos de Agentes de IA (I)

Marcelo do Nascimento Oliveira Soares

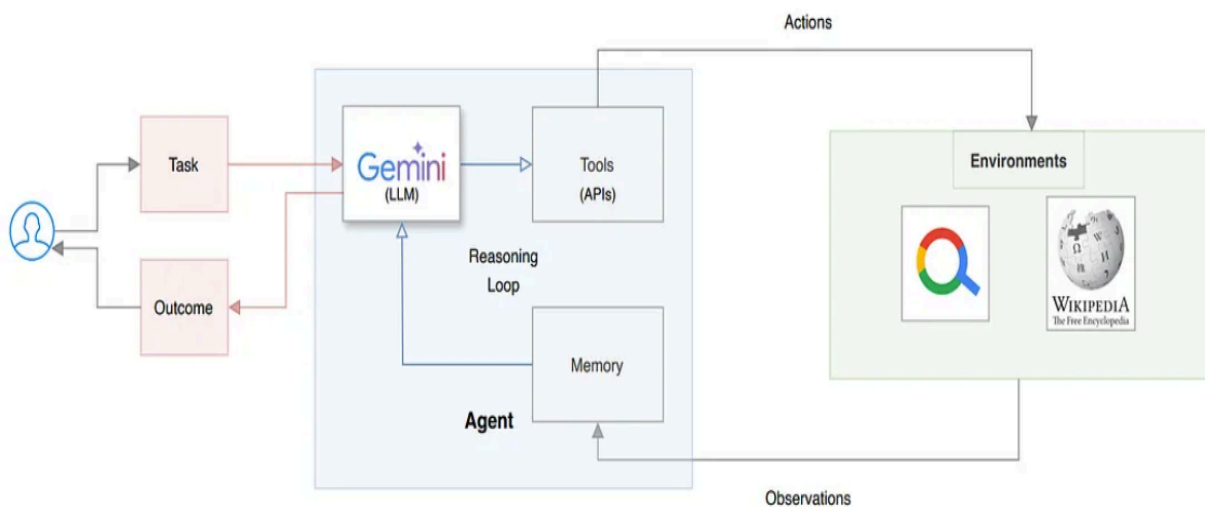
Descrição da atividade:

Analisando os temas abordados, creio que o fato de maior relevância é que qualquer uso superficial das LLMs está obsoleto. A cada dia novas abordagens são criadas, novas formas de usar esses modelos de linguagem são implementadas, testadas e aprovadas. É o que Andrew, no segundo vídeo menciona, os avanços tecnológicos atualmente e, segundo sua previsão, no futuro, não se devem mais ao desenvolvimento de hardware ou de novas estruturas que tragam mais performance, mas ao uso do que já foi conquistado e, pensando em uso, o que realmente faz diferença são as abordagens para resolver diferentes problemas.

Dessa forma, considerando problemas clássicos como: os LLMs são limitados aos dados de treinamento, tornando-os antiquados diante de novas descobertas e de fatos recém publicados. Diante desse problema, os RAGs surgiram como uma solução, agora além dos dados de treinamento as LLMs podem acessar novas fontes e até mesmo se especializar nelas.

Além da abordagem que considera o uso dos RAGs, o primeiro vídeo aborda outra metodologia, nesse método o LLM é implementado com intuito de raciocinar e em seguida agir. A inteligência artificial costumava separar a tomada de decisões da execução, agora, nessa metodologia a tomada de decisões e a execução de processos funcionam de forma cíclica. O fluxo do método ReACT pode ser descrito em entrada, raciocínio, ação com o ambiente externo, observação e memória, feedback loop e saída:

- Entrada: O LLM age como o “cérebro” do agente e, partindo de uma “entrada”, esse modelo de linguagem define para o agente quais e como as tasks serão abordadas.
- Raciocínio: O LLM analisa a task e a fragmenta, para resolver fragmento por fragmento. Com a fragmentação da task o agente pode decidir quais ações tomar com base nas ferramentas e informações disponíveis.
- Ação com o ambiente externo: o agente busca em ambientes externos mais contexto.
- Observação e memória: O agente observa os resultados e salva as informações mais relevantes.
- Feedback Loop: Cada passo anterior é repetido até que todas as subtasks sejam contempladas pelas informações salvas pelo agente na memória.
- Saída: O agente devolve ao usuário as informações obtidas e refinadas a cada processo.



- Essa imagem, disponibilizada no primeiro link de referência, descreve o fluxo que descrevi acima.

Dificuldades:

Acho que uma dificuldade vai ser implementar e entender melhor a arquitetura de um ReACT Agent.

Pensando nessa atividade, o vídeo 2 foi mais difícil de compreender. No começo pensei que o Andrew estava falando em outra língua além de inglês, mas no final ele só falava mais rápido e com um sotaque que eu demorei mais pra entender.

Conclusões:

Acredito que o mais relevante no conteúdo é o avanço tecnológico possibilitado pela pesquisa e pelo desenvolvimento de novas abordagens que sempre possuem um contraste em relação às abordagens anteriores. Pensando em contraste, para finalizar este relatório, deixo uma tabela abaixo exemplificando a ruptura de paradigma nos aspectos que considero mais importantes:

Mudança	LLM	Ai Agents
base de conhecimento	Limitada ao banco de dados utilizado no treinamento do modelo. Para ter mais contexto é necessário apelar a outras técnicas.	Por conta da volatilidade do seu fluxo, que já conta com a possibilidade de acessar dados externos ao modelo, a base de conhecimento é limitada apenas pela disponibilidade de dados online.
Memória	Capaz de armazenar o histórico dos inputs	Capaz de armazenar o histórico de inputs e utilizá-los no aprendizado.

Autonomia	“Raciocínio” limitado, capaz de abstrair padrões em sua base de dados e produzir saídas limitadas às informações armazenadas no modelo.	Possui um raciocínio interno, capaz de produzir feedbacks e construir resultados melhores após cada iteração.
Tarefas	Responde perguntas considerando o contexto	Pode dividir a tarefa em subtarefas e decidir como irá chegar ao resultado esperado
Ferramentas	Limitada	Pode usar diversas ferramentas disponíveis.

Referências:

<https://medium.com/google-cloud/building-react-agents-from-scratch-a-hands-on-guide-using-gemini-ffe4621d90ae>

<https://www.youtube.com/watch?v=F8NKVhkZZWI>

<https://www.youtube.com/watch?v=KrRD7r7y7NY>