

Agentes

Fabián Villena

Introducción

Los agentes autónomos se han reconocido como un acercamiento prometedor para alcanzar la inteligencia artificial general en donde se espera resolver tareas a través de acciones y planificaciones auto guiadas.

Antes, los agentes se concebían como funciones altamente restringidas al dominio en donde fueron ajustadas, pero esto es alejado del aprendizaje humano, en donde se aprende desde una variedad de ambientes.

Agentes

Un agente es una entidad artificial capaz de percibir su ambiente a través de sensores, tomar decisiones y posteriormente proceder utilizando actuadores.

El concepto de agente se originó en la filosofía en donde se describen como entidades que poseen deseos, creencias, intenciones y la habilidad de actuar.

- Agentes simbólicos: Siguen una serie de reglas para resolver una tarea.
- Agentes reactivos: Pueden percibir y reaccionar sobre el ambiente.
- Basados en aprendizaje reforzado: Aprenden a través de la interacción.
- Con transferencia de aprendizaje: Migración de aprendizaje desde otras tareas.
- **Basados en LLMs**

Agentes en inteligencia artificial

La definición de agentes se trasladó hacia la ciencia de la computación con la intención de dotar a computadores con la habilidad de entender los intereses de los usuarios y de manera autónoma actuar en su nombre.

Al avanzar la IA, el término se utiliza para denotar entidades que muestran un comportamiento inteligente y poseen cualidades como la autonomía, reactividad, proactividad y habilidades sociales.

Se cree que los agentes son el punto de inflexión para alcanzar la Inteligencia Artificial General.

Un camino hacia la IAG

Para lograr reales avances para entender el lenguaje, debemos lograr que los modelos enfrenten el mundo a diferentes niveles, los cuales se los llama World Scopes:

1. **Corpus:** Recopilar, preprocesar y anotar documentos para estudiar cómo representarlos.
2. **Internet:** La utilización de datos no estructurados, no etiquetados, multi-dominio al nivel de recopilar todo lo que la humanidad ha escrito, mejora nuestras representaciones.
3. **Percepción:** Se incorporan nuevas modalidades de datos.
4. **Personificación:** Se incluye el actuar sobre el mundo, seguir instrucciones y entender las implicancias físicas del lenguaje.
5. **Social:** Se añaden interacciones para entender intenciones, emociones y los efectos del lenguaje sobre otros.

Agentes autónomos basados en LLMs

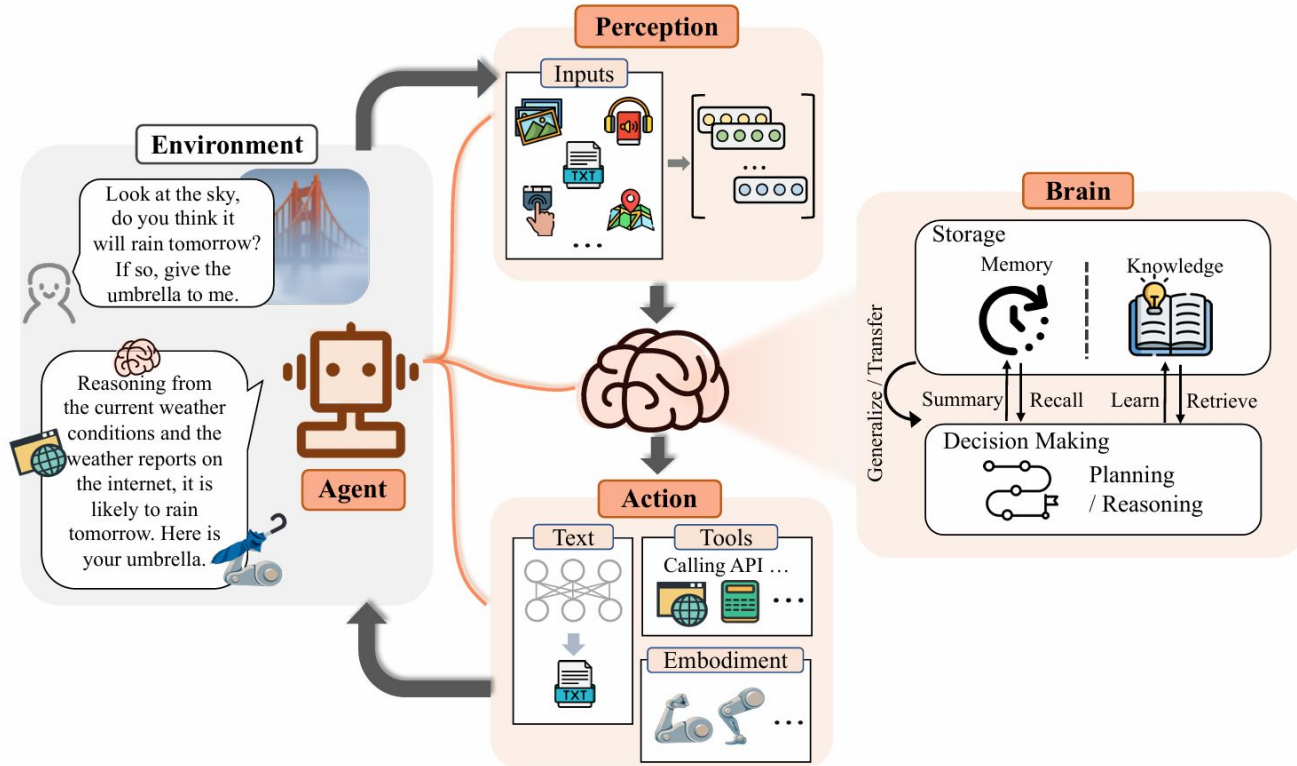
Al elevar los LLMs al estatus de agentes autónomos y equiparlos con herramientas para expandir su espacio de percepción y acción, podríamos potencialmente llegar al tercer y cuarto nivel de WS.

Además, estos agentes podrían ejecutar tareas más complejas a través de la cooperación y competencia, emergiendo así fenómenos sociales que podrían ayudarnos a llegar al quinto nivel.

Sociedad donde interactúan agentes y humanos



Cerebro, percepción y acción



Cerebro

Este módulo está basado en un LLM y sirve como el núcleo central de un agente.

Después de recibir información desde el módulo de percepción, el agente primero recupera conocimiento desde su almacenamiento de conocimiento y memoria. Esto ayuda al modelo a tomar decisiones más informadas. Además el modelo puede memorizar cosas nuevas.

Las facetas que demuestra el cerebro son:

- Interacción en lenguaje natural
- Conocimiento
- Memoria
- Razonamiento y planificación
- Generalización

Percepción

Es crucial que los LLM reciban información desde varias fuentes y modalidades dado que este espacio perceptual ampliado ayuda a los agentes a entender de mejor manera su ambiente, realizar decisiones informadas y destacar en muchas tareas. Esta información debe ser traspasada hacia el cerebro.

Algunas entradas perceptuales pueden ser:

- Texto
- Visual
- Auditiva
- Olfativa, táctil, mapas tridimensionales, etc.

Acción

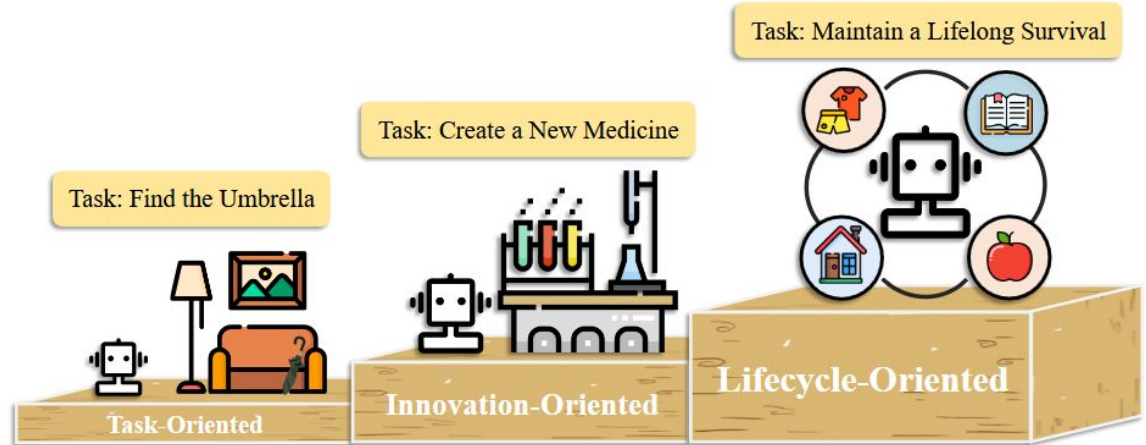
Después de percibir la información, integrarla, analizarla, razonar sobre ella y tomar decisiones, se debe actuar sobre el ambiente. El agente toma las acciones que salen del cerebro y las ejecuta sobre el ambiente.

Algunas acciones que el agente puede realizar son:

- Texto
- Utilización de herramientas
- Personificar
 - Observar
 - Manipular
 - Navegar

Habilidades de un agente único

Son los agentes más valorados y más prevalentes dado que son capaces de mejorar la eficiencia en alguna tarea y disminuir la carga de trabajo de un usuario.



El potencial de múltiples agentes

La falta de interacciones sociales entre agentes genera que el paradigma de un sólo agente restrinja el potencial que tienen los agentes.

La división del trabajo puede llevar a que los agentes alcancen su rendimiento óptimo al enfocar el trabajo de cada uno en resolver una sola tarea.

