

Problema 1 de 4

SID 2425q2

Febrero/Marzo 2025

El termino agente (*agents*) está siendo recientemente muy popular en la comunidad de la IA. Esto es debido al uso del concepto combinado con modelos grandes de lenguaje (LLMs) para obtener software autónomo capaz de realizar tareas en nombre de un usuario o de mostrar un comportamiento competente en entornos complejos.

Elegid una (1) propuesta reciente, como por ejemplo:

- Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023, October). Generative agents: Interactive simulacra of human behavior. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (pp. 1-22): <https://dl.acm.org/doi/pdf/10.1145/3586183.3606763>
- Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). React: Synergizing reasoning and acting in language models. arXiv preprint arXiv:2210.03629: <https://arxiv.org/pdf/2210.03629.pdf>
- Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2024). Reflexion: Language agents with verbal reinforcement learning. Advances in Neural Information Processing Systems, 36: https://proceedings.neurips.cc/paper_files/paper/2023/file/1b44b878bb782e6954cd888628510e90-Paper-Conference.pdf
- Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., & Anandkumar, A. (2023). Voyager: An open-ended embodied agent with large language models. arXiv preprint arXiv:2305.16291: <https://arxiv.org/pdf/2305.16291.pdf>
- Nakajima, Yohei (2023). Task-driven autonomous agent¹ / BabyAGI: <https://github.com/yoheinakajima/babyagi>

¹<https://yoheinakajima.com/task-driven-autonomous-agent-utilizing-gpt-4-pinecone-and-langchain-for-diverse-applications/>

Podéis encontrar más propuestas en páginas web como Google Scholar o arXiv.org, buscando por palabras clave como *agentic systems*, *llm agents* o *agent architecture*.

Analizad la propuesta en términos de lo que habéis visto en teoría de SID y responded a las siguientes preguntas:

1. Arquitectura (*6 puntos*)
 - (a) ¿Hay algún elemento (métrica, objetivo, forma de evaluación) descrito en la propuesta en el que los autores definan la racionalidad desde el punto de vista del agente propuesto? (*2 puntos*)
 - (b) ¿Cómo describiríais la arquitectura de agente que se propone? ¿Qué componentes se describen, cómo se motiva que se incluyan y cómo se relacionan entre sí? (*2 puntos*)
 - (c) ¿A qué arquitectura, de las vistas en teoría, diríais que se ajustaría más esta propuesta? ¿Por qué? (*2 puntos*)
2. Entorno (*4 puntos*)
 - (a) ¿En qué entorno o entornos pueden operar los agentes propuestos? (*2 puntos*)
 - (b) Caracteriza dicho(s) entorno(s) en función de las propiedades vistas en teoría. (*2 puntos*)

Se dará **2 puntos extra** sobre los 10 a aquellos grupos que basen su respuesta al ejercicio en alguna propuesta que no sea una de las listadas anteriormente, con fecha de publicación en 2024 o 2025. Por favor, referenciad adecuadamente el documento, repositorio o artículo que contiene la propuesta con la que habéis trabajado.

En la entrega tiene que ser un documento breve (.pdf) que incluya una solución al ejercicio propuesto, además de una sección donde expliquéis cómo os habéis repartido las tareas entre los autores de la entrega.

La entrega y corrección del problema se hará en grupos de como máximo 3 personas (4 personas, de manera excepcional) y contará un 1/4 de la nota de problemas. Deberéis entregar vuestra solución antes del final del domingo 9 de marzo en el espacio que se habilitará en el Racó.