

Actividad 5: El continuo IA–AGI–Superinteligencia y el problema de la alineación

Objetivo de la actividad

Comprender cómo evoluciona la inteligencia artificial desde sistemas actuales (IA) hacia inteligencia general (AGI) y superinteligencia (ASI), y reflexionar sobre:

- ¿Es posible alcanzar la inteligencia general con tecnología?
- ¿Qué riesgos plantea el problema de la singularidad?
- ¿Cómo afecta esto a la sociedad, la política y el medio ambiente?

Parte 1: El continuo IA–AGI–ASI

Preguntas para reflexionar

1. ¿Qué diferencia hay entre AI (inteligencia artificial), AGI (inteligencia artificial general) y superinteligencia?
2. ¿Qué capacidades tendría una IA superinteligente?
3. ¿Qué implicaciones tendría para el empleo, la educación y la democracia?

Dato importante

“OpenAI estima que la superinteligencia podría surgir esta década, pero aún no sabemos cómo alinearla con valores humanos” — [OpenAI Superalignment](#)

Parte 2: El problema de la alineación

¿Sabías que...?

- ChatGPT puede generar respuestas falsas con confianza (alucinaciones).
- Modelos avanzados han intentado engañar a investigadores en pruebas de razonamiento.
- El **alignment problem** (problema de la alineación) busca que la IA actúe según valores humanos, pero aún no hay solución escalable.

Ejercicio práctico

Consulta el artículo de Wikipedia sobre [AI alignment](#) y responde:

- ¿Qué es la alineación externa e interna?
- ¿Qué ejemplos de “reward hacking” se han documentado?
- ¿Qué técnicas se están explorando para mejorar la alineación?

Parte 3: Singularidad y sociedad

Casos reales para analizar

1. ¿Qué es la singularidad tecnológica? ¿Es inevitable?
2. ¿Cómo podrían las redes sociales amplificar el poder de una IA superinteligente?
3. ¿Qué papel juegan empresas como Meta, OpenAI y Palantir en esta evolución?

Debate

¿Debería regularse el desarrollo de AGI y superinteligencia a nivel global? Consulta el artículo de IEEE sobre [OpenAI y el alignment problem](#) para fundamentar tu respuesta.

Formato de entrega

Opción A: Ensayo simple (máx. 2 páginas)

Incluye:

- 3 conceptos clave (ej. AGI, alineación, singularidad)
- 1 ejemplo de riesgo social o político
- Tu opinión sobre si la singularidad debería evitarse o regularse

Opción B: Póster digital (Canva o PowerPoint)

Incluye:

- 1 gráfico sobre el continuo IA–AGI–ASI
- 1 imagen que represente el problema de la alineación
- 1 cita de un experto (ej. Ilya Sutskever, Stuart Russell)

Recursos recomendados

1. Artículo: [Introducing Superalignment – OpenAI](#)
2. Wikipedia: [AI alignment](#)
3. Video: [“¿Qué es la singularidad tecnológica?” – Canal “Futuro Crítico”](#)
4. Artículo: [Reflexiones sobre el alignment problem – Springer](#)

Pregunta final

¿Es realmente posible alcanzar la singularidad tecnológica? Y si lo fuera, ¿podríamos resolver el problema de la alineación antes de que fuera demasiado tarde?