Gobernanza y ética en la era de los agentes de IA

La metáfora del coche

- LLM ≠ agente. Un LLM es el motor: genera potencia lingüística y responde.
- El agente es el coche completo: añade memoria, percepción, planificación, ...
- Sin chasis, sensores ni frenos, el motor por sí solo no te llevará a ningún sitio.
- Conducir un coche sin gobernanza equivale a lanzar un agente sin reglas.



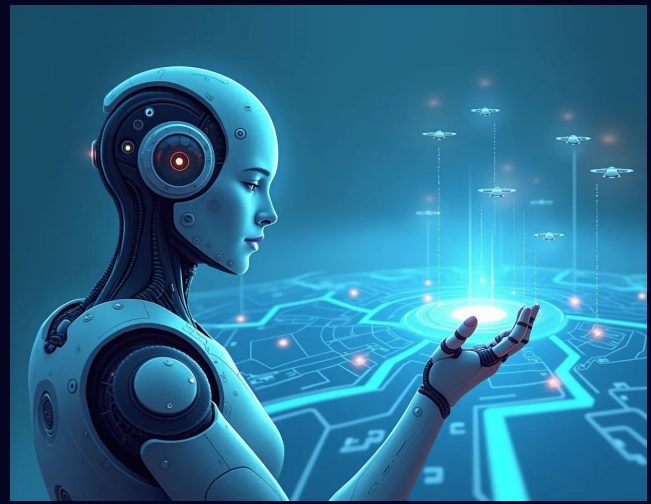


Definición de agente

“Sistema que percibe, planifica y ejecuta acciones de forma autónoma para alcanzar objetivos, pudiendo aprender de la experiencia y delegar tareas.”

IMPORTANTE:

Clave: autonomía ≠ ausencia de supervisión.



Panorama estratégico: oportunidad, ciclo de vida y arquitectura

- Oportunidad de negocio
- Reducción de costes ≈ 30 % en operaciones de atención al cliente gracias a generative AI (McKinsey).
- Ejemplo de retorno rápido: Verizon declara un +40 % en ventas tras usar un asistente Gemini para sus 28 000 agentes (Reuters).

¿Cuándo elegir un agente?

- Tarea repetitiva pero con variabilidad (ej. reclamaciones, onboarding).
- Necesidad de orquestar varias herramientas/APIs.



Por qué hablar hoy de gobernanza

- Normativas empiezan a exigir trazabilidad, seguridad y explicabilidad.
  - AI Act EU - Primer marco jurídico integral del mundo para la inteligencia artificial - La AI Act convierte la IA europea en una autopista reglada: lo inaceptable se prohíbe, lo arriesgado se controla y lo cotidiano circula con confianza.
  - NIST RMF - Revisión de la cadena de suministro, que no entre nada corrupto al sistema, que no salga nada dañino del sistema.
- 2025: los grandes actores apuestan por estándares abiertos para agentes (p. ej. Model Context Protocol) y memoria estructurada. Por lo que es necesario crear un marco de gobernanza y ética.
  - Microsoft wants AI 'agents' to work together and remember things - Microsoft apuesta por agentes web que interactúan entre ellos.
  - AI Agents: Why We Should Strategize on Governance - Sin visibilidad y trazabilidad, los agentes serán actores sin dueño ni control.

Los tres retos de los próximos meses

Reto 1 – Memoria

- De simple chat a conversaciones continuas.
- Ejemplo: alergia a frutos secos → el agente debe recordarlo siempre que sugiera recetas.
- Tendencias: “structured retrieval augmentation”, resúmenes vectoriales, memoria a corto y largo plazo.



Reto 1 – Memoria

- Memory for agents - Sin memoria, un agente es como un pez dorado: nada rápido, pero a los tres segundos olvida por qué
- Microsoft wants AI 'agents' to work together and remember things - Memoria eficiente: para evitar los altos costes de cómputo
- Agent Memory - La memoria es el puente que convierte cada conversación en un capítulo, no en un reinicio



Reto 1 – Memoria

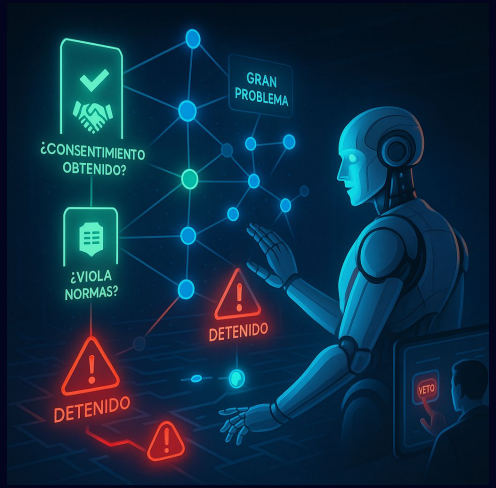
Ángulo de gobernanza y ética:

- Propiedad y caducidad del dato: permitir al usuario ver, exportar y borrar recuerdos.
- Mitigación de sesgos con metadatos de procedencia.
- Auditoría: registrar qué recuerdo disparó qué decisión



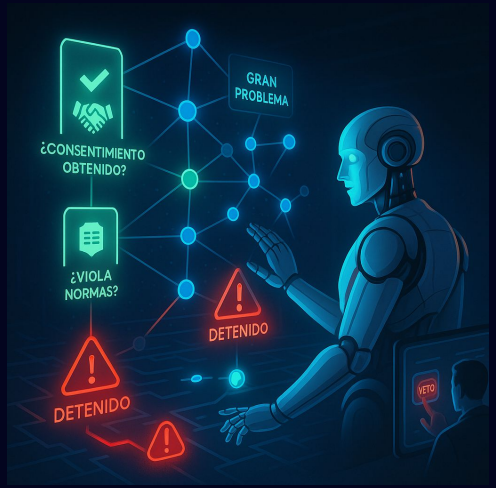
Reto 2 – Planificación robusta

- Dividir problemas → sub-tareas → evaluaciones intermedias.
- Técnicas: ReAct, meta-razonamiento, guardian agents, grafos de tareas.



Reto 2 – Planificación robusta

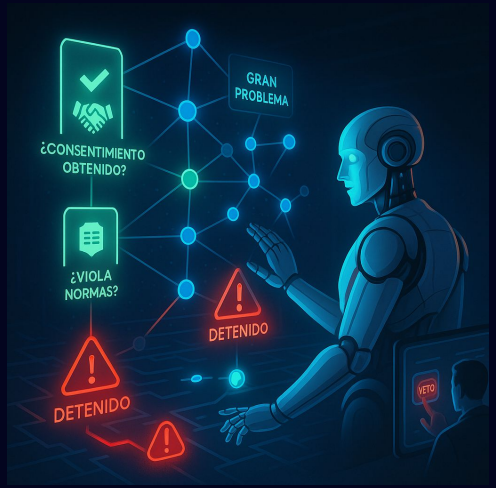
The evolving ethics and governance landscape of agentic AI - Dar control a un LLM es como cederle el volante sin frenos automáticos y un conductor responsable, la solución se convierte en peligro



Reto 2 – Planificación robusta

Ángulo de gobernanza y ética:

- Inyectar checkpoints éticos (¿hay consentimiento? ¿viola normas?).
- Estrategias de “fail-safe”: detenerse cuando el plan requiera pasos inseguros.
- Transparencia: mostrar al usuario el plan a alto nivel; permitir veto.



Reto 3 – UX para agentes

- Pasamos de hacer clic a delegar intención. UX debe informar qué hará el agente.
- Diseños emergentes: chat enriquecido, ambient agents, paneles de seguridad.



Reto 3 – UX para agentes

- Hello AI Agents: Goodbye UI Design, RIP Accessibility - Cuando tu agente conduzca, nadie verá el salpicadero: diseña para inteligencias que interactúan sin mirar la pantalla
- The Ethics Of AI In UX: Designing Transparent And Fair Experiences - En la UX de agentes, la transparencia es el diseño: tu copiloto digital debe explicar cada maniobra para que el usuario detecte (y corrija) sesgos antes de que el sistema actúe



Reto 3 – UX para agentes

Ángulo de gobernanza y ética:

- Human-in-the-Loop: controles claros para aprobar, parar o revertir.
- Accesibilidad: multimodalidad (voz, braille, gestos) → no dejar colectivos fuera.
- Explicabilidad visual: por qué se eligió este paso y no otro.



Conclusiones

- El motor ya existe; el coche no. Gobernar memoria, planificación y UX decide si será un vehículo seguro o un riesgo público.
- Ética = calidad: agentes que respetan al usuario son agentes más útiles y adoptables.
- Llamado a la acción: incorpora estos principios en prototipos desde el primer momento.

