Claude en serio

Taller práctico de automatización y primer agente

SÁBADO · 18:30 – 20:30

"Un agente no es un humano más rápido.
Es una regla más rápida. Y las reglas,
sin humano que las revise, se rompen en silencio."

VI Studio

En 120 minutos sales con

- Tu **primer agente** funcionando sobre un caso real tuyo
- Las **3 reglas inviolables** de ese agente escritas
- Claro el **mapa de autonomía**: qué puede hacer solo y qué no
- Los **5 guardrails** que evitan que la líe
- Saber cuándo un agente **te ahorra horas** y cuándo te mete en un lío

De brazo a cabeza

SESIÓN 3 (CONECTORES)	HOY (AGENTES)
"Lee mi Gmail y respóndeme"	"Lee, **decide** y ejecuta solo"
Tú dictas cada paso	Claude itera **sin pedir permiso** por cada cosa
Un brazo que alcanza tus datos	Una **cabeza** que decide qué hacer con ellos

> *Ayer le pusimos brazos a Claude. Hoy le ponemos cabeza.*

Qué es un agente — los 3 ingredientes

1. **Memoria** — quién eres, cómo trabajas (tu system prompt de la sesión 1)
2. **Herramientas** — los conectores de ayer (Gmail, Drive, Calendar)
3. **Autonomía** — decide y ejecuta varios pasos seguidos, sin parar a preguntarte

> *Memoria + herramientas + autonomía = agente. Quítale el 3 y vuelve a ser un chat.*

El caso de Lourdes — del conector al agente

Ayer quedó pendiente: comparar importes del ERP contra el Excel de control y cantar las diferencias del cierre.

```
Cada día 5 del mes:
1. Abre el ERP y el Excel de control (conectores)
2. Compara importe a importe
3. Marca las diferencias > 50 €
4. Genera una tabla "a revisar" con causa probable
5. Te la deja en el Drive y te avisa
```

> *Eso ya no es "léelo y respóndeme". Es un agente:* ***decide qué comparar, ejecuta y vuelve con el resultado.***

El mapa de autonomía — 4 niveles

NIVEL	QUÉ HACE	RIESGO
1 · Leer	Lee y te resume	✅ libre
2 · Proponer	Prepara el borrador, **tú envías**	✅ libre
3 · Ejecutar	Actúa solo, **con límites duros**	⚠️ guardrails obligatorios
4 · Decidir por ti	Decide cosas que afectan a personas	🚫 zona prohibida

> ***Hoy construimos nivel 1 y 2.*** *El 3 lo aprendes a blindar. El 4 ni se toca.*

7×

un "equipo de agentes" gasta
unas 7 veces más que un chat normal

Una sesión compleja puede quemar 1,3 M de tokens en minutos. Empieza simple.

Reto · 60 segundos

Escribe las **3 reglas inviolables** de tu agente.
Lo que **nunca** debe hacer, pase lo que pase.

> "Nunca envía un email sin que yo lo lea. Nunca toca importes > 1.000 €. Nunca usa datos de un cliente sin marcar."

⏱️ Cuenta atrás. Lo lees a tu mesa.

30 min · primer mini-entregable

Pega en el chat común:

> "Mi agente va a ___ , es de nivel ___ (1 leer / 2 proponer), y su regla nº1 inviolable es ___"

Si tu agente es nivel 3 a la primera, bájalo a 2. Se escala con el tiempo, no el primer día.

Skills — para que el agente sea repetible

Una **Skill** es una instrucción guardada que Claude carga **solo cuando la llamas**.

SIN SKILL	CON SKILL
Reexplicas el flujo cada vez	Escribes `/cierre-mensual` y arranca
Claude improvisa	Sigue tu lista de pasos fija
Resultado distinto cada día	Mismo proceso, mismo formato

> *Es la diferencia entre explicarle a un becario cada mañana y darle un manual.*

Cowork — el modo "déjale trabajar"

CHAT	COWORK
Una pregunta → una respuesta	Una tarea → minutos de trabajo
Tú marcas el ritmo	Claude itera solo y te reporta
Cada paso lo ves	Te avisa cuando termina

> *Para tareas que antes te ocupaban una tarde entera. Disponible en la app de escritorio con plan de pago.*

Ejercicio · 60 minutos

Construye **tu primer agente** (nivel 1 o 2) sobre un caso real:

1. Elige una tarea repetitiva tuya (triaje, cierre, clasificación de leads)
2. Escribe sus instrucciones: qué lee, qué decide, qué entrega
3. Conéctale **1 herramienta** de ayer
4. Dale las **3 reglas inviolables** del reto
5. Ejecútalo con un caso real y **verifica el resultado a mano**

⏱️ Yo paseo entre mesas. Foco: que **funcione**, aunque sea pequeño.

90 min · segundo mini-entregable

Enseña tu agente al compañero/a:

- ¿Qué tarea automatiza?
- ¿De qué nivel es?
- ¿Qué hizo bien y qué tuviste que corregir?

90 segundos por persona. Cuenta atrás.

Cápsula legal · 10 minutos · nivel 3

Si tu agente **ejecuta solo** (nivel 3), necesitas los 5 guardrails:

1. **Registro de todo** — cada acción con fecha, qué entró, qué hizo
2. **Límites duros** — importes, frecuencia, a quién puede escribir. Por escrito
3. **Botón de paro** — que lo frene en 5 segundos
4. **Revisión humana** — cada X acciones, alguien mira
5. **Plan de incidente** — qué haces si la lía: quién avisa, quién responde

Art. 22 RGPD — la línea roja

> ***No puede haber decisiones 100 % automáticas que afecten a una persona*** *sin un humano en el bucle.*

- ❌ Un agente que **rechaza** a un candidato solo
- ❌ Un agente que **deniega** un crédito solo
- ❌ Un agente que **sanciona** a un empleado solo
- ✅ Un agente que **prepara** y un humano que decide

> *Si tu caso roza esto: hoy aprendes el marco, el lunes llamas a tu DPO.*

AI Act — alto riesgo

Estos usos son **alto riesgo** desde agosto 2026 → documentación + supervisión humana obligatoria:

- RR.HH. (selección, evaluación)
- Scoring crediticio
- Sanidad y justicia

> *Para Oronis / financiero: cuidado con cualquier agente que puntúe o decida sobre personas. El resto de lo que montáis hoy es riesgo mínimo.*

Las 4 trampas más comunes

TRAMPA	DEFENSA
"Que pague facturas solo desde el día 1"	Empieza nivel 1 → escala tras 30 días sin incidente
Quemar 80 € en tokens con "equipos de agentes"	Agente simple, no equipos, salvo caso justificado
Darle el OK por encima	Mini-checklist + revisión del resultado siempre
Meterle datos sensibles sin DPA	Frenazo legal — repasa la cápsula

La IA no te quita la responsabilidad

de lo que decide el agente.

Te la multiplica.

Cierre del bloque · entregable visible

✅ Tu primer agente nivel 1-2 funcionando
✅ Sus 3 reglas inviolables escritas
✅ Sabes los 5 guardrails para escalar a nivel 3
✅ Tienes claro dónde está la línea roja legal

Mañana: llevamos esto a **tu propio ordenador** con Claude Code — control total, privacidad y modelos que viven en tu máquina.

Fin del sábado

Mañana 10:00 · puntuales