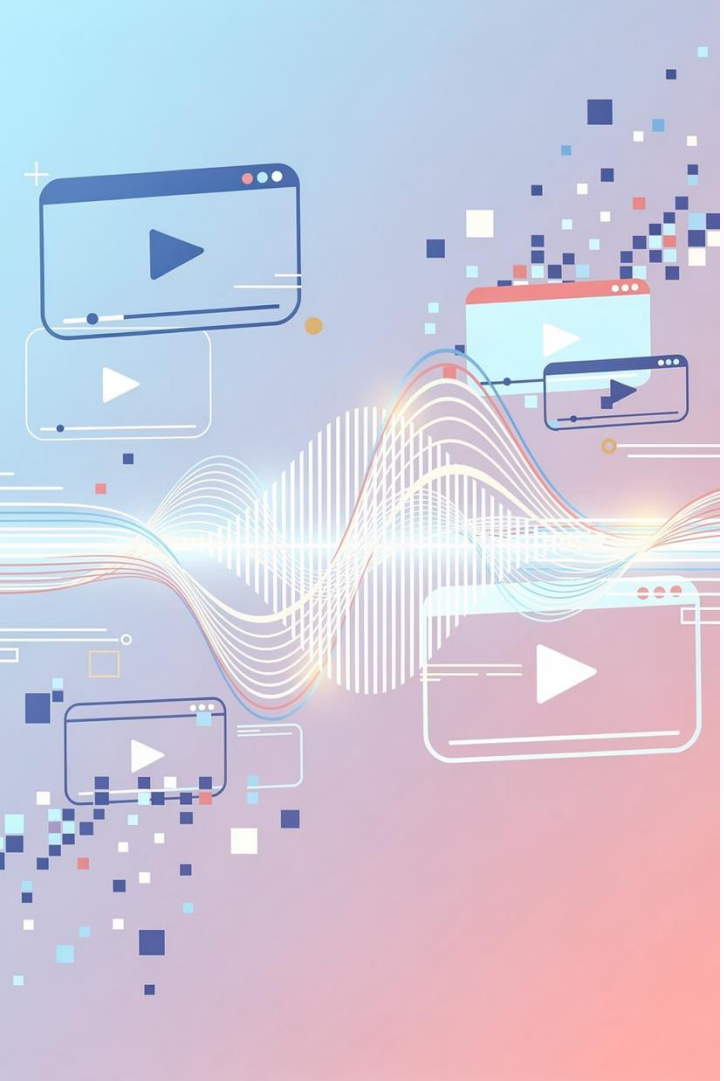


IA Generativa: Imagen · Audio · Vídeo

IA GENERATIVA: COMPETENCIAS DIGITALES, SEGURIDAD Y APLICACIONES PRÁCTICAS

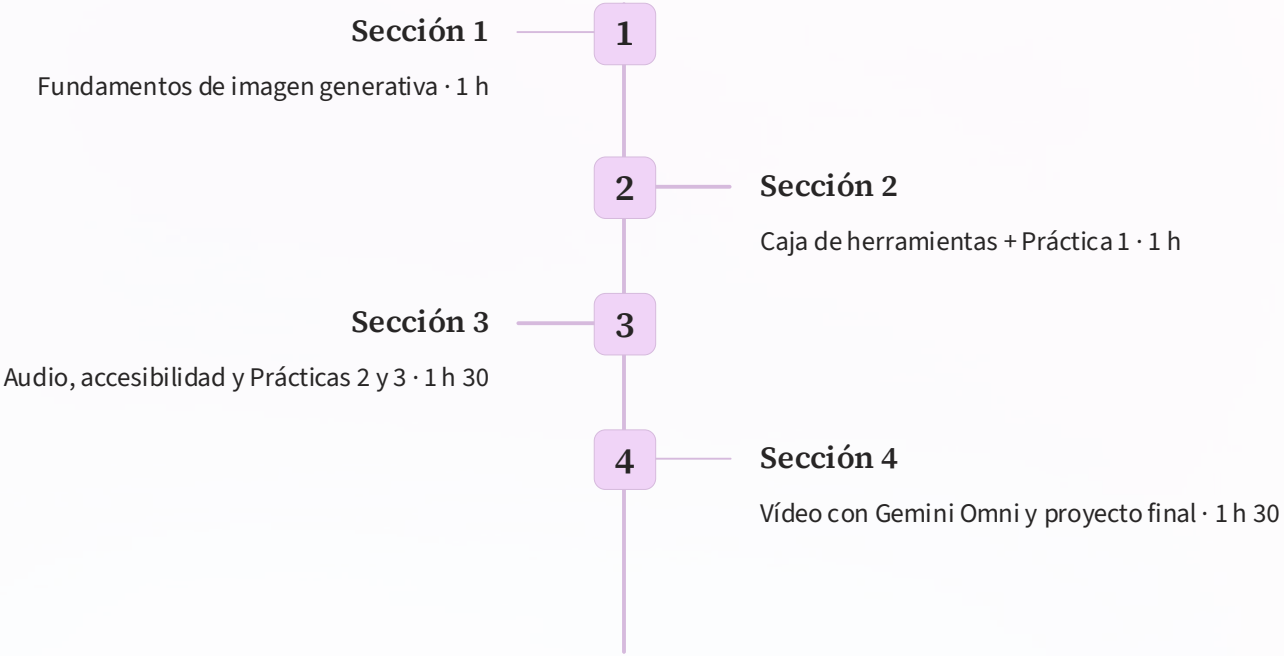
De la idea al vídeo, sin perder el criterio. Imagen generativa, audio, accesibilidad y vídeo con Gemini Omni, Veo 3.1 y compañía.

5 horas · 4 secciones · 4 prácticas · Junio 2026

Centro de Investigación Operativa · Universidad Miguel Hernández

Cómo recorreremos estas 5 horas

Bienvenida y plan del curso



Dos pausas (15 + 10 min). Cuatro prácticas en grupos de 3-4 personas. Cada grupo elige al inicio un **servicio municipal** sobre el que trabajará todo el día.

SECCIÓN 1

Generación de Imágenes

Conceptos, oportunidades y desafíos

01

imágenes?

02

Modelos de difusión: de ruido a realidad

03

Oportunidades reales en la administración local

04

Licencias, sesgos y texto

05

La anatomía del prompt + iteración

No busca: crea

La IA generativa no consulta bancos de imágenes ni hace búsquedas. **Genera píxeles nuevos** a partir de patrones aprendidos durante el entrenamiento.

- **Text-to-Image**: texto entra, imagen sale.
- Es un **copiloto creativo**: tú describes, ella propone, tú afinas, ella regenera.
- La primera imagen casi nunca es la definitiva.

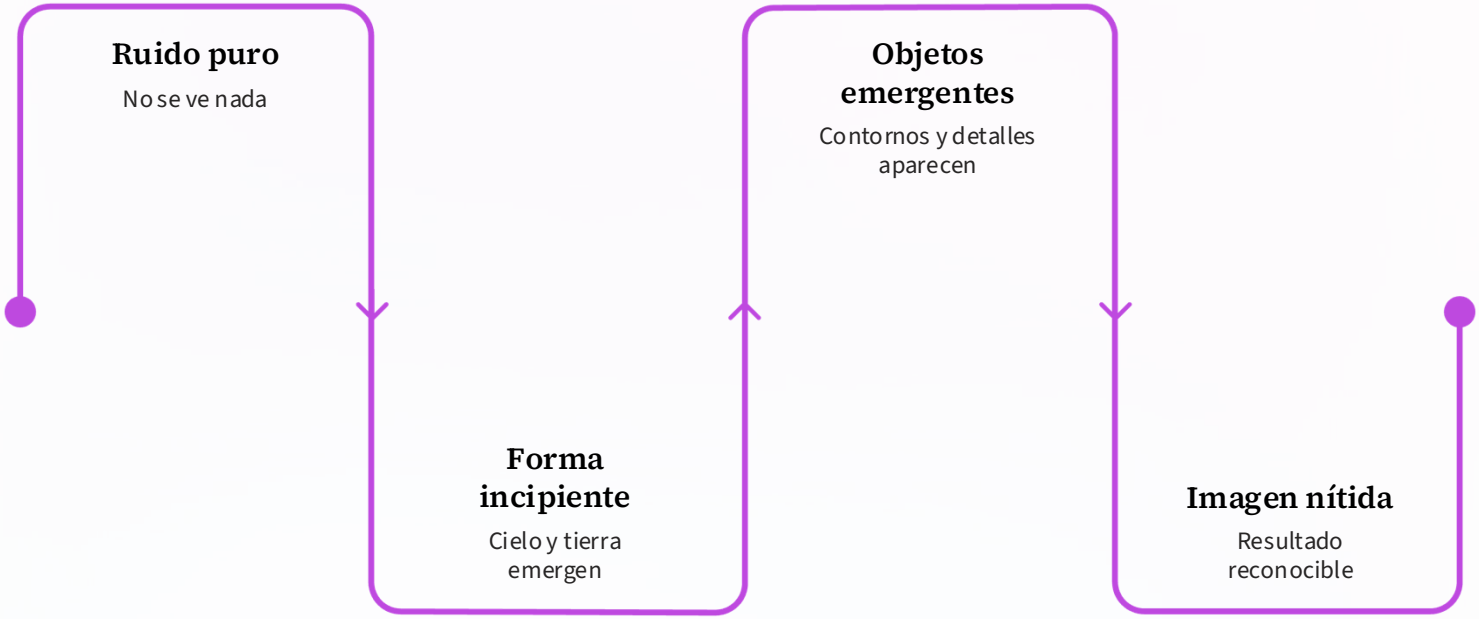
> Para fotografía real documentada → banco de imágenes. Para conceptos y simulaciones → IA generativa.



La magia de los modelos de difusión

De ruido puro a imagen nítida, paso a paso

La IA aprende "rompiendo" imágenes con ruido y memorizando cómo reconstruirlas. Al recibir un prompt, parte del ruido y va organizándolo progresivamente.



Cambiar una sola palabra del prompt redirige todo el proceso. Por eso dos generaciones con el mismo texto nunca son idénticas.



Oportunidades para el Sector Público

Agilidad, ahorro y personalización

Concienciación ciudadana

Piezas inclusivas a coste de minutos: reciclaje, igualdad, prevención.

Redes y efemérides

Respuesta inmediata para fechas y eventos locales.

Simulaciones urbanísticas

Previsualizar una calle peatonalizada antes de la obra.

Patrimonio histórico

Recrear edificios desaparecidos para educación y turismo.

Desafíos: Licencias, Sesgos y Texto

Y un añadido legal de 2026

⚖️ Licencias

Cada modelo tiene condiciones distintas. **Adobe Firefly** es la opción más segura comercialmente.

🛡️ Sesgos heredados

Si no pides diversidad explícita, no aparece. Hay que forzarla en el prompt.

🔤 Texto en imagen

Ideogram y Gemini lo manejan bien; Midjourney lo destroza.

⚠ **AÑADIDO 2026 — Reglamento de IA de la UE**
Obliga a etiquetar todo contenido sintético. Google marca con **SynthID** cada imagen y vídeo generados automáticamente.

La anatomía del prompt perfecto

Cinco ingredientes que separan amateur de profesional

1

1 · Sujeto

Quién o qué protagoniza la imagen

2

2 · Acción

Verbo: qué hace el sujeto

3

3 · Entorno

Dónde, contexto, hora del día

4

4 · Estilo

Fotorrealista, ilustración, render 3D

5

5 · Composición

Plano, lente, iluminación

> *"Familia diversa paseando por una plaza peatonalizada al atardecer, estilo fotorrealista, gran angular con luz cálida."*

Iterar el prompt: de amateur a profesional

El mismo "cartel de reciclaje", tres versiones

V1 · Amateur

"Cartel sobre reciclaje"

Resultado: imagen de stock genérica, sin contexto local ni personalidad.

V2 · Con anatomía

"Vecinos reciclando vidrio en una plaza con bancos y árboles, ilustración plana en verdes y azules."

Usable, pero todos los vecinos se parecen.

V3 · Profesional

Añade diversidad explícita (edades, silla de ruedas, rasgos variados) y reserva espacio para el lema. **Listo para imprimir.**

Lección clave: Iterar es la norma, no la excepción. Reservad **5 minutos por pieza**, no 30 segundos.

SECCIÓN 2

La Caja de Herramientas

Los modelos relevantes en junio de 2026

01

LMArena, Copilot y Adobe Firefly

02

Ideogram, Nano Banana (Gemini), Midjourney

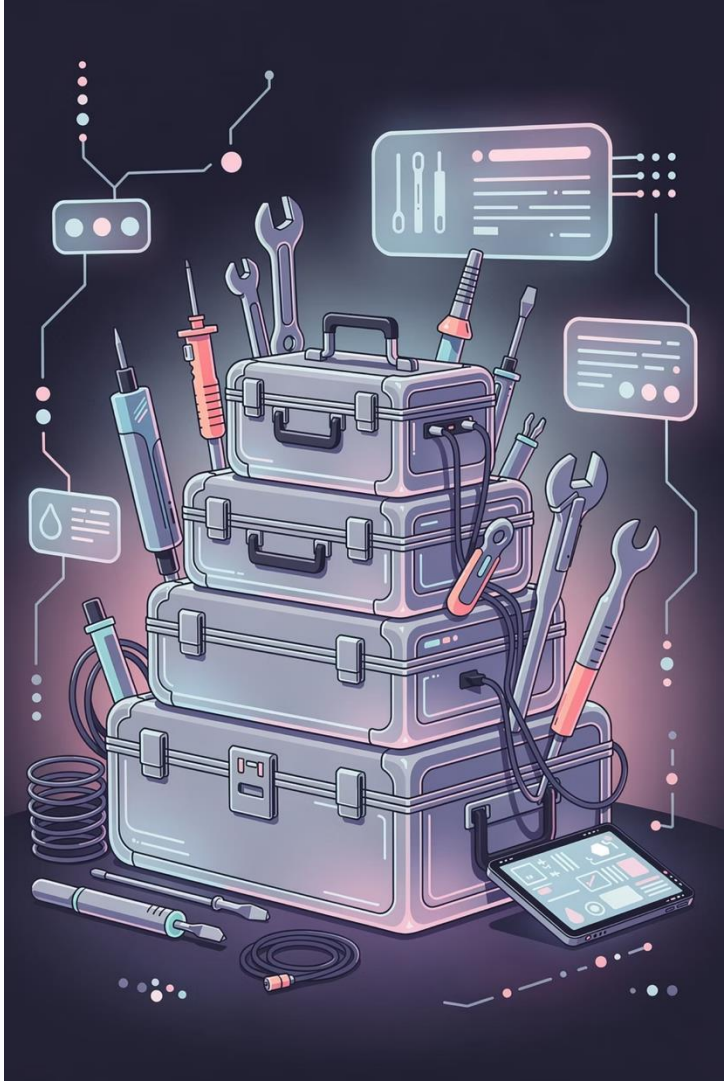
03

Meta AI y Flux Kontext

04

PRÁCTICA 1: imagen para tu día a día

El mercado cambia cada mes. La selección es de junio 2026; el **criterio que aprendéis sirve siempre**.



LMArena · Copilot · Adobe Firefly

El punto de entrada y los dos básicos

LMArena

Web gratuita para **comparar modelos a ciegas**. Muestra dos generaciones sin revelar cuál es cuál: tú votas. Imprescindible antes de pagar suscripciones.

Microsoft Copilot

Punto de partida si ya tenéis Office 365. **Integrado en Word y PowerPoint**. Ideal para ilustración corporativa del día a día.

Adobe Firefly

La apuesta **comercialmente segura**: entrenada solo con material licenciado. Para cuando la responsabilidad legal es innegociable.

Ideogram · Nano Banana · Midjourney

Los especialistas



Ideogram AI

El rey del **texto legible en imagen**. Carteles, infografías, logos provisionales. Primera opción para comunicación ciudadana con lemas.



Nano Banana (Gemini)

Edición conversacional: "quita el coche", "pon más gente". Excelente coherencia de personajes entre generaciones.



Midjourney

Referencia estética: **fotorrealismo y calidad cinematográfica**. Talón de Aquiles: el texto sigue saliéndole irregular.

A menudo, una pieza final combina varias de estas tres herramientas.

Midjourney · La belleza por encima de todo

La referencia estética del mercado

Versión actual

V7 (por defecto) y **V8.1 alpha** desde abril 2026: HD 2K nativo, 5× más rápido y mejor texto.

Sin plan gratuito

La calidad cuesta. **10 $/mes (Basic)** hasta **120 $/mes (Mega)**. Para uso profesional realista: 30 $/mes (Standard).

Donde brilla

Portadas de revista municipal, postales turísticas, recreación de patrimonio histórico, ilustraciones cinematográficas.

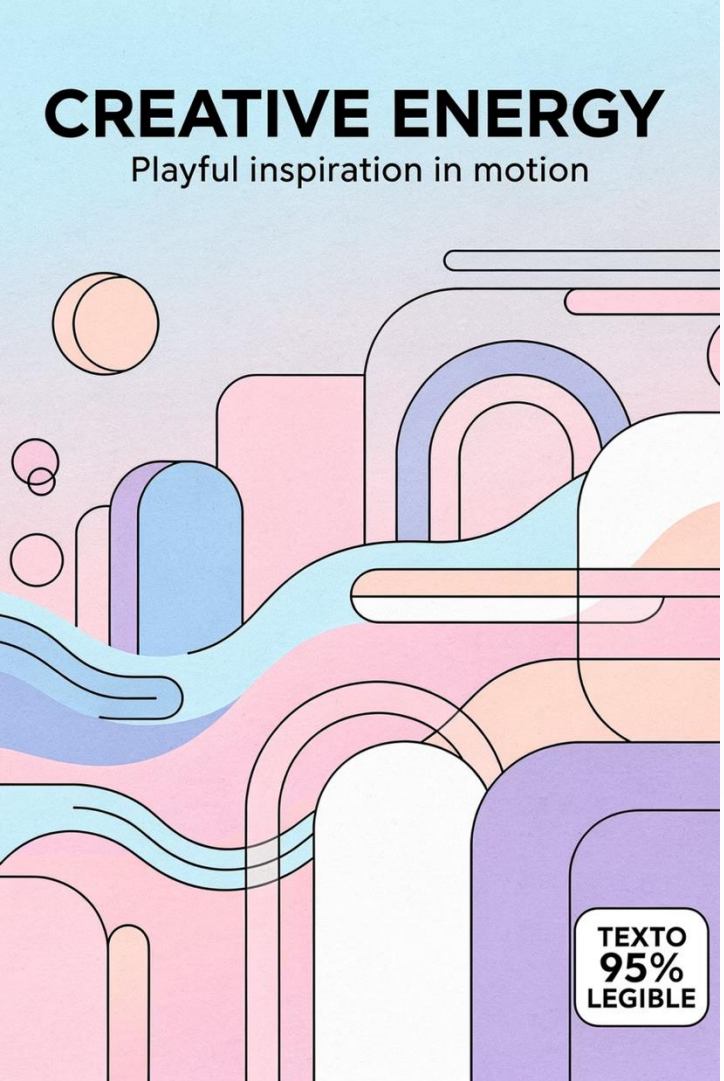
Donde flojea

El texto dentro de imagen sigue siendo su talón de Aquiles, aunque V8.1 mejora mucho.

Novedad 2026

Ya genera **vídeo desde imagen** (V1, hasta 21 segundos).





Ideogram · El rey del texto legible

90-

Modelo actual

Ideogram 3.0 (refinado durante 2026). Especialista absoluto en tipografía legible.

Plan gratuito real

~10 generaciones diarias para pruebas y ejercicios de aula (verifica la web, ajustan cuotas).

Precios

Plus 7 $/mes, Pro 42 $/mes anual. El "premium" más barato del mercado.

Características clave

Magic Prompt (mejora tu prompt automáticamente), **Canvas** (edición por capas), **Magic Fill** (inpainting selectivo), entrenamiento de marca propia.

Su sitio

cualquier pieza con texto — carteles ciudadanos, infografías, logos provisionales, miniaturas con titular.

Comparativa rápida: lo que Midjourney resuelve en 15 intentos con texto, Ideogram lo hace al segundo.

Nano Banana Pro · El editor inteligente

Gemini 3 Pro Image: la opción "premium" para piezas importantes

Qué es

la versión flagship del editor de imágenes de Google, lanzada en noviembre 2025 sobre Gemini 3 Pro.

Capacidades

hasta **4K de resolución**, mezcla **hasta 14 imágenes** en una sola composición, mantiene **hasta 5 personas consistentes** entre planos distintos.

Texto multilingüe

genera texto legible en castellano, valencià y otras lenguas. Localiza y traduce texto **dentro** de la propia imagen.

Edición conversacional avanzada

ángulo de cámara, iluminación, color grading, recomposición — todo con lenguaje natural.

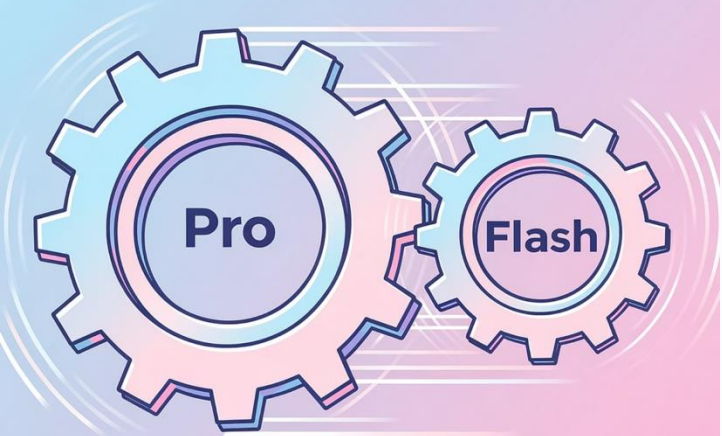
SynthID por defecto

cada imagen lleva marca de agua de Google.

Acceso

cuota gratuita limitada en la app Gemini; cuota completa con **Google AI Plus o Pro (~20 $/mes)**.





Nano Banana 2 · Velocidad Flash con potencia Pro

Para el flujo de trabajo diario: rápido, multilingüe y ya en tu Workspace

Qué es

Gemini 3.1 Flash Image, lanzado en marzo 2026. Combina las **capacidades de Pro con la velocidad de Flash**.

Dónde aparece

app Gemini, **Google Workspace**, **NotebookLM**, Google Ads, AI Mode en Búsqueda. **Probablemente ya está en tu ayuntamiento** si tenéis Workspace.

Su ventaja

edición e iteración casi instantáneas. Iterar 10 versiones sin esperar.

Hereda de Pro

SynthID, texto multilingüe legible, comprensión del mundo real desde Google Search (datos, geografía, eventos actuales).

Cuándo elegir cuál

Pro

→ pieza final (cartel de campaña, portada de folleto, render hero).

2 (Flash)

→ todo lo demás: drafts rápidos, edición conversacional, exploración, iteración en grupo.

Recraft · La cuarta pieza que falta

Diseño gráfico, no fotografía: la herramienta de marca corporativa

Qué es

el único generador pensado para **diseño gráfico** en lugar de imagen fotográfica. Output como **SVG, PNG, JPEG, TIFF y Lottie**.

Lienzo estilo Figma

trabajas en un canvas que crece, no en prompts sueltos. Permite combinar varios modelos en el mismo proyecto (V4 propio + GPT-4o + Imagen 4 + Flux + Ideogram + HiDream).

Plan gratuito generoso

30 créditos al día. Suficiente para uso profesional ligero, ideal para aula.

Donde brilla

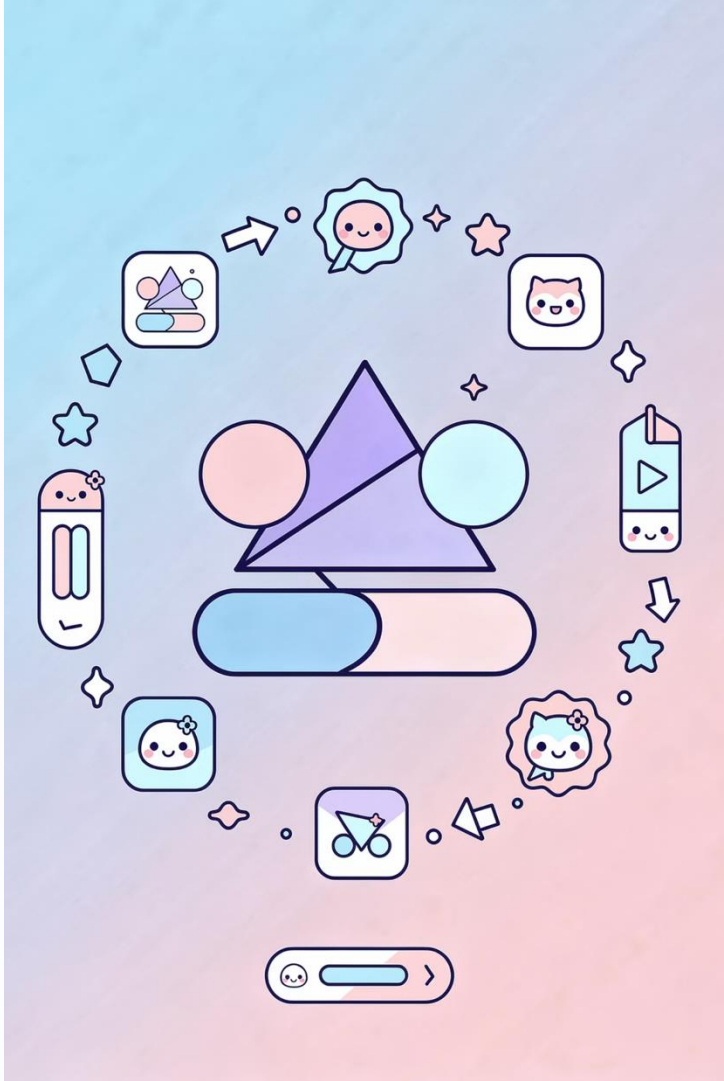
logos provisionales, iconos para infografías municipales, ilustraciones de identidad corporativa, banners con texto, mockups, vectorización (raster → SVG).

Característica única

estilos de marca personalizados. Entrenas el modelo con la paleta y tipografía del ayuntamiento y mantiene la coherencia entre piezas.

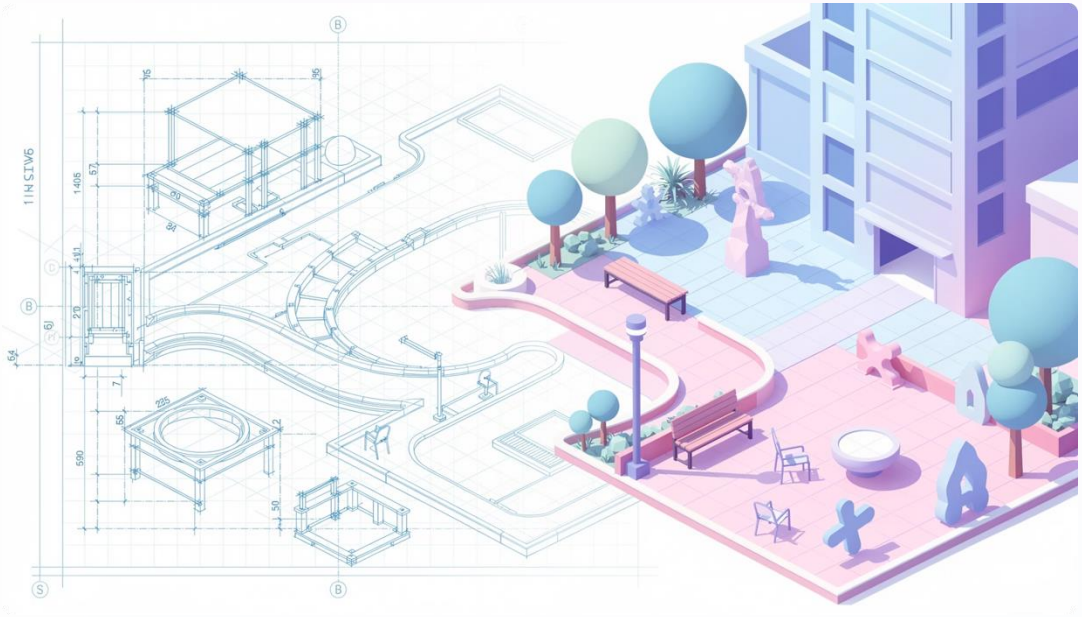
Limitación honesta

para fotorrealismo de personas no es la mejor; para marca y diseño sí.



Meta AI · Flux Kontext

Dos alternativas con nicho claro



Meta AI

Rápida y fotorrealista. Útil para **visualizar escenarios de crisis** en formación interna (Protección Civil, bomberos).

⚠ Una imagen de incendio realista mal etiquetada genera pánico. Etiquetado obligatorio.

Flux Kontext

De boceto o plano a **render fotorrealista**. La joya silenciosa para urbanismo: enseñar a una asociación de vecinos cómo quedará la plaza tras la obra.

PRÁCTICA 1

Transformando Necesidades Diarias

≈ 40 min · Grupos de 3-4 personas

1

Cabecera para PPT

Imagen conceptual con Firefly o Copilot.

2

Diagrama técnico

Apoyo visual para informe con Nano Banana o Ideogram.

3

Cartel ciudadano con texto

Ideogram, sin dudarlo.

4

Contenido para RR. SS.

Copilot o Midjourney según el tono deseado.

Reparto: 5 min organización · 30 min producción · 5 min selección. Cada grupo presenta **2 piezas** (no 4) en 1 min cada una.



SECCIÓN 3

De la Imagen al Movimiento

Audio, accesibilidad y nuevos roles

A partir de aquí cambia el oficio: aparece el guion, el sonido y la **responsabilidad sobre el discurso**.

01

Ecosistema audiovisual y nuevos roles

02

El prompt multimedia: cámara, sonido, ética

03

PRÁCTICA 2: diseñar el prompt base

04

Audio, accesibilidad y música

05

PRÁCTICA 3: locución y música

El Ecosistema Audiovisual y los Nuevos Roles

De "creador" a cuatro roles simultáneos

Guionista

Decide qué se cuenta y en qué orden. **Primera decisión, más responsable.**

Director/a de escena

Planos, ritmo, montaje. **Cómo se cuenta visualmente.**

Revisor/a ética

Sesgos, inclusión, honestidad. Es un papel, no necesariamente una persona aparte.

Supervisor/a legal

Licencias, derechos de imagen, datos personales, etiquetado AI Act.

La IA no inventa el mensaje del ayuntamiento; **ejecuta el que tú le des.**

El Prompt Multimedia: el Detalle Importa

Seis dimensiones que un prompt de imagen no necesitaba



Crear contenido audiovisual requiere especificar dimensiones que la imagen estática no necesitaba.

⚠ **Las restricciones suelen ser más importantes que las inclusiones.** Indicar qué NO debe aparecer (logos, menores, marcas) evita los errores más costosos.

PRÁCTICA 2

Diseñar el Prompt Base

≈ 25 min · En papel y boli

Cada grupo elige un servicio municipal y redacta **tres prompts en papel**, sin tocar herramientas todavía.

Prompt 1 · Locución

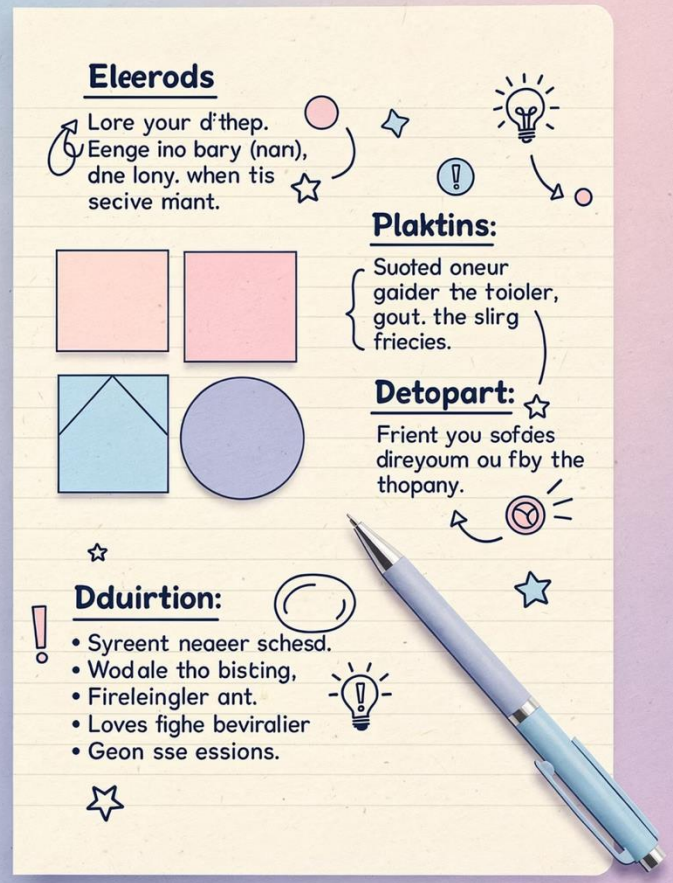
Texto para sistema TTS con el tono correcto. Se prueba en la Práctica 3.

Prompt 2 · Vídeo

Plano, duración, ambiente, elementos visuales. Se prueba en la Práctica 4.

Prompt 3 · Revisión ética

Lista de preguntas antes de publicar. Se queda con el grupo como herramienta de trabajo permanente.



SECCIÓN 3

ElevenLabs · Voz institucional

Síntesis y clonación de voz con calidad humana

- **Text-to-Speech** en más de 70 idiomas, incluidos castellano, valencià y catalán.
- **Tres usos**: locución general, doblaje multilingüe automático y clonación de una voz concreta.
- **Clonación profesional**: requiere 30 min mínimo de audio limpio del hablante **y consentimiento expreso por escrito** de esa persona. No oral, no implícito.
- **Plan gratuito = NO comercial**. Para uso institucional, mínimo plan Starter (~5 $/mes); plan Creator (~22 $/mes) si quieres clonación profesional.
- **Recomendación**: para la "voz oficial" del ayuntamiento, elige una voz sintética genérica y úsala siempre la misma. Más fácil de mantener, de etiquetar y de defender.





Suno · Música y jingles

De un prompt a una canción completa en 60 segundos

- **Modelo v5.5** (marzo 2026): vocales, instrumentación y arreglo en un solo paso. Pistas de hasta 8+ minutos.
- **Genera letra, melodía y voz** desde una descripción de texto: *"jingle pop alegre de 30 s sobre empadronamiento, voz femenina, tono cercano"*.
- **Plan gratuito = NO comercial**. Para uso municipal, mínimo plan **Pro (~10 $/mes)**: licencia comercial y ~500 canciones/mes.
- **Plan Premier (~30 $/mes)**: incluye **Suno Studio**, una DAW nativa para editar pistas y separar stems.
- **Atención 2026**: panorama legal en movimiento. Warner Music ya firmó acuerdo de licencia con Suno; los modelos van a migrar a catálogo licenciado durante 2026. **Documenta siempre la licencia y la versión del modelo** que usaste; guarda el recibo.

Audio Generativo, Accesibilidad y Música

Tres usos, tres cautelas

Voz institucional (ElevenLabs)

Locución, doblaje multilingüe y clonación de voz. **Solo con consentimiento expreso por escrito.**

Música y jingles (Suno)

Jingles para campañas en segundos. La versión gratuita **NO sirve** para uso institucional. Revisar siempre la licencia comercial.

Accesibilidad

Versiones en audio de documentos administrativos para discapacidad visual, mayores y ciudadanía extranjera. **La idea más rentable del curso.**

PRÁCTICA 3

Locución y Música

≈ 30 min · Material listo para emitir

Entregable 1 · Aviso de corte de agua

Formal

Para web y comunicaciones oficiales.

Cercana

Para redes sociales del ayuntamiento.

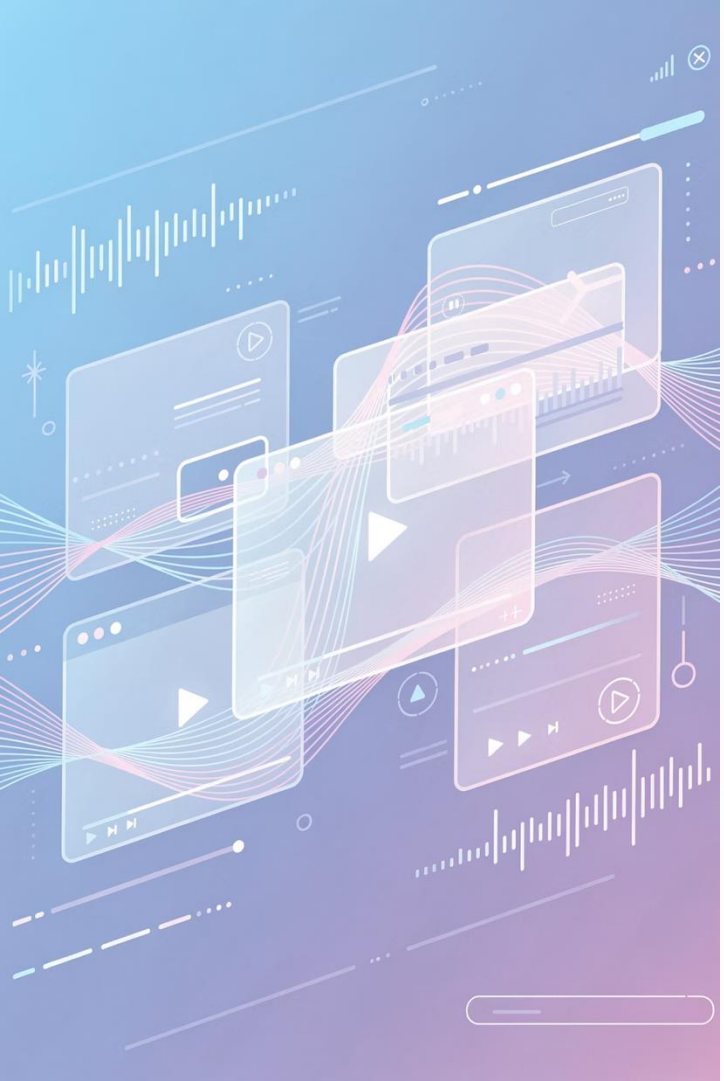
Telegráfica

Para centralita telefónica automática.

Entregable 2 · Jingle de empadronamiento

Jingle pop de **30 segundos** sobre empadronamiento. Pegadizo, claro, **licencia revisada**.

⚠ Si el jingle se usa de verdad, no salir del aula sin **documentar la licencia comercial**.



SECCIÓN 4

Vídeo Generativo y Proyecto Final

El estado del arte en junio 2026

Lo que veréis sobre Gemini Omni se anunció hace tres semanas en **Google I/O 2026**. Lo importante no son los nombres: es el patrón que se repite.

01

Gemini Omni vs Veo 3.1

02

Alternativas: Kling, Hailuo, Runway, Seedance 2.0

ĆĊ

PRÁCTICA 4: taller de creación de vídeo

ĆÇ

Proyecto integrado + transparencia

Gemini Omni vs. Veo 3.1

La gran novedad de Google I/O 2026 — y por qué Veo no muere

Gemini Omni

App Gemini · Flow · YouTube

- Entrada multimodal: texto + imagen + audio + vídeo
- Edición conversacional turno a turno
- Clips de ~10 s con audio sincronizado
- Coherencia de personajes entre planos
- Marca de agua SynthID por defecto

Veo 3.1

Vertex AI · Google Flow · Perfil profesional

- Realismo cinematográfico hasta 4K
- Audio nativo de alta calidad
- Producción de mayor duración

Regla práctica: Para borradores y redes → **Omni**.
Para piezas publicables y largas → **Veo 3.1**.

Alternativas y Aplicaciones

Cuatro herramientas más a tener en el radar

Kling

Realismo físico convincente: gravedad, líquidos. Para piezas que deben **parecer rodadas**.

Hailuo

Optimizado para vertical y corta duración. Su sitio: **Reels y TikTok institucional**.

Runway

La herramienta de edición más madura. Útil si en el equipo hay **perfil audiovisual**.

Seedance 2.0

El rival serio de Omni y Veo (ByteDance). **Calidad cinematográfica** con multi-referencia.

Aplicaciones municipales: previsualizar obras · campañas de limpieza · simular espacios públicos · portales de transparencia.

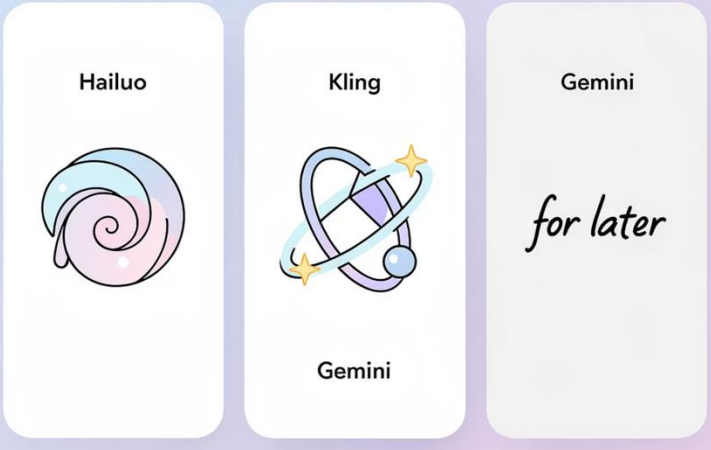
Vídeo IA gratuito · panorama junio 2026

Qué te dan, qué te quitan

Herramienta	Plan gratis	Resolución / duración	Limitaciones
Kling 3.0	66 créditos/día (no acumulan)	720p · 5-10 s	Marca de agua, cola lenta
Runway Gen-4	125 créditos one-time al registrarse	720p · ~25 s totales	Es una prueba, no un plan
Hailuo 2.3	Créditos diarios de lanzamiento	1080p · 6-10 s	Cuota diaria, prioridad baja
Gemini Omni Flash	Incluido en Gemini gratis	~10 s con audio	Marca SynthID por defecto
Veo 3.1	Acceso parcial vía Gemini	Hasta 4K en pago	Restringido sin suscripción

Aviso: Sora ya no juega. Cerrada durante 2026; sus usuarios se han ido sobre todo a Kling.





Cuál usar?

La pareja Hailuo + Kling, con Gemini Omni como comodín

- **Hailuo 2.3 · primero**. El mejor "wow" gratuito ahora mismo: 1080p, 10 segundos, motion coherence sorprendente. MiniMax está regalando créditos diarios durante el lanzamiento del modelo.
- **Kling Free · segundo**. 66 créditos diarios **predecibles**: cada alumno sabe qué tiene. Ideal para ideas que se generan a la vez.
- **Gemini Omni · comodín**.
- **Runway**. Excelente calidad, pero los 125 créditos son **one-time** al registrarse. Resérvalo para hacer una interfaz profesional, no para producir en aula.

Avisos clave y receta

Cuatro cautelas y una propuesta para la Práctica 4

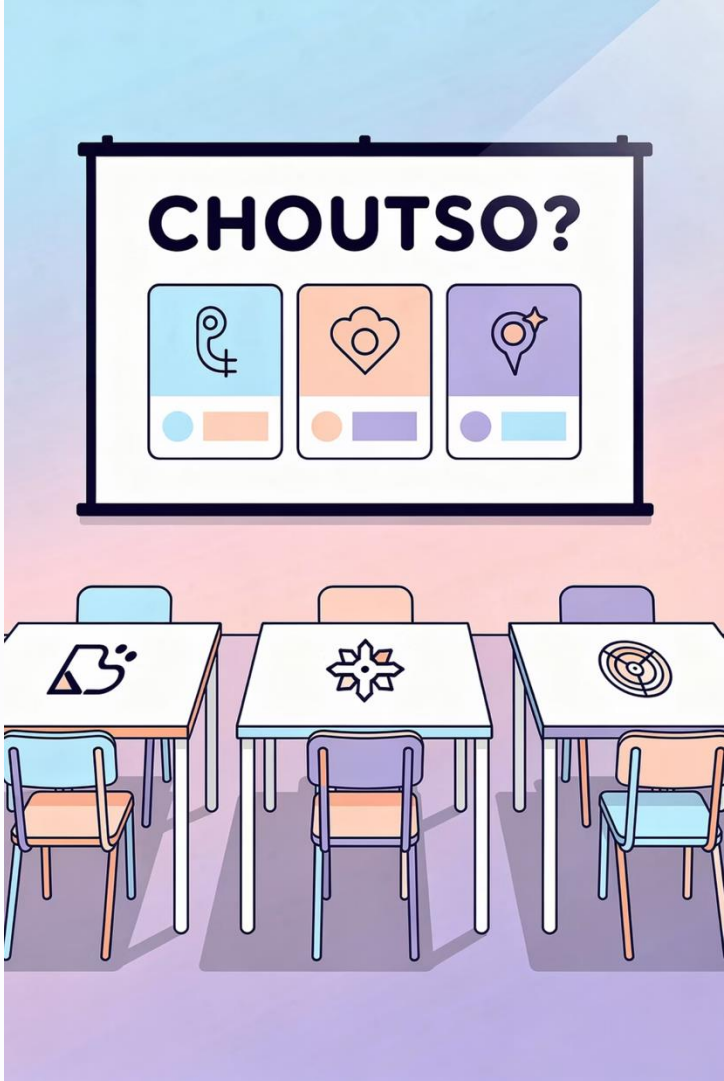
Avisos antes de empezar:

- **Marca de agua = sí** en casi todos los tiers gratuitos. Ejercicios de clase, sí. Publicación en redes oficiales, no.
- **Uso comercial = casi siempre NO** en planes gratuitos. Cualquier pieza con marca municipal hacia el exterior exige plan de pago documentado.
- **Créditos diarios > créditos one-time** para un curso: cada día se arranca con cuota nueva, no hay sorpresas a media mañana.
- **Sora cerrada**: si alguien la menciona, redirigir a Kling o Hailuo.

Receta para la Práctica 4 (5 grupos):

- 2 grupos prueban **Hailuo 2.3** → motion coherence.
- 2 grupos prueban **Kling Free** → calidad física.
- 1 grupo prueba **Gemini Omni** → "cero fricción Workspace".

Al final, los 5 clips lado a lado en pantalla y la pregunta clave: **¿cuál firmarías con el logo del ayuntamiento mañana?**





PRÁCTICA 4

Taller de Creación de Vídeo

· Pieza pequeña pero acabada

1

Vídeo de 8-10 s

Sobre turismo o movilidad sostenible. Borrador con Omni; si hace falta calidad, Veo.

2

Animar una foto

Del archivo municipal: imagen fija → clip de unos segundos.

3

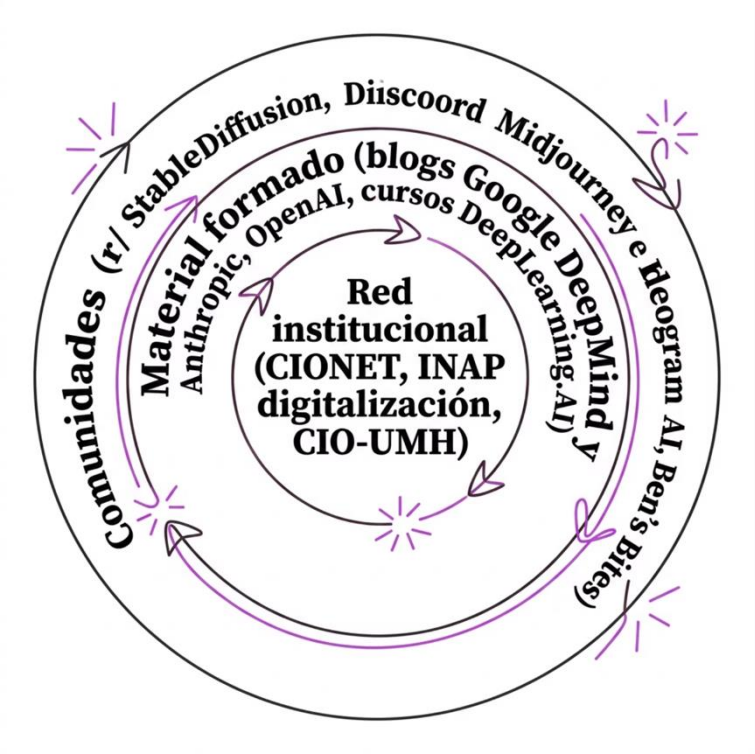
Refinar prompt institucional

Claro e inclusivo, partiendo del prompt de la Práctica 2.

Reparto: 5 min organización · 30 min producción · 10 min revisión por pares. Antes de cerrar: cada grupo identifica **qué etiquetar como contenido sintético**.

Para Seguir Aprendiendo

El día después del curso



Tres niveles de recursos

🌐 Comunidades

r/StableDiffusion, Discord de Midjourney e Ideogram, boletines como The Rundown AI y Ben's Bites.

📚 Material formativo

Blogs de Google DeepMind y Anthropic, anuncios de OpenAI, cursos cortos de DeepLearning.AI.

🏛️ Red institucional

CIONET, grupos del INAP sobre digitalización, CIO-UMH.

✅ **Regla práctica**: 15 minutos a la semana > 1 curso al año. Montad un grupo interno de práctica mensual en el ayuntamiento.

Proyecto Integrado y Transparencia

El flujo completo y la regla de oro



Objetivo | Guion | Voz y Música | Vídeo | Edición

Checklist de ética antes de publicar

- ¿Hay rostros reales o menores identificables?
- ¿Tenemos derechos de uso de música, voz e imagen?
- ¿Es accesible? (subtítulos, audiodescripción, lenguaje claro)

REGLA DE ORO — Transparencia

La ciudadanía tiene derecho a saber qué contenido es sintético. **Etiquétalo siempre.**
Coste de etiquetar: **cero**. Coste de no etiquetar: que la próxima vez que digáis algo verdadero, no os crean.