



MANUAL DEL PARTICIPANTE

# Marketing con IA Multimodal

Una jornada de 4 horas para crear contenido profesional con IA

Jornada de 4 horas · Taller presencial  
CONFA · Marketing con IA Multimodal  
Edición 2026

# Bienvenida

Esta es tu jornada de 4 horas para construir un flujo profesional de creación de contenido multimodal con IA. Saldrás con piezas reales producidas por ti: copy, imagen, video y audio, integradas en una mini campaña que podrías publicar mañana mismo.

## CLAVE Lo que vamos a lograr

Pasarás de un brief a una pieza terminada (estática + animada + locutada) usando ingeniería de prompts multimodal, ElevenLabs (ElevenCreative) y FloraFauna.ai. El énfasis está en técnica de prompt, no en clic-clic.

## Lo que asumimos que ya manejas

- Estructura básica de prompt: rol, contexto, objetivo, formato.
- Técnicas estándar como zero-shot, few-shot y chain-of-thought.
- Uso de un LLM conversacional (ChatGPT, Claude, Gemini o Manus).

## Lo que aprenderás aquí (y no se enseña en cursos básicos)

- Anatomía de un prompt visual de 7 capas para imagen y video.
- Cómo encadenar copy → imagen → video → voz manteniendo coherencia.
- Direccionar emoción, ritmo y SFX con Audio Tags v3 de ElevenLabs.
- Construir flujos nodales en FloraFauna y en ElevenLabs Flows.

## Accesos para hoy

Herramienta	Acceso
ElevenLabs (acceso Confa)	<a href="https://try.elevenlabs.io/confa">https://try.elevenlabs.io/confa</a>
FloraFauna.ai	<a href="https://flora.ai/">https://flora.ai/</a>
Manus (invitación)	<a href="https://manus.im/invitation/KKHKSDN9TIY5TXU">https://manus.im/invitation/KKHKSDN9TIY5TXU</a>

## ATENCIÓN Antes de empezar

Crea cuenta en ElevenLabs por el link de Confa antes del Bloque B y en Flora antes del Bloque C. Ten a mano un brief real (cliente, producto, objetivo, audiencia, tono) — lo usaremos toda la jornada.

## AGENDA · 4 HORAS

# Cronograma de la jornada

Hora	Bloque
0:00 – 1:15	Bloque A — Ingeniería de prompts multimodal
1:15 – 2:30	Bloque B — ElevenLabs / ElevenCreative a fondo
2:30 – 2:45	Receso 15 min
2:45 – 3:45	Bloque C — FloraFauna.ai a fondo
3:45 – 4:00	Cierre · Flujo end-to-end y entrega

## BLOQUE A · 75 MIN

# Ingeniería de prompts multimodal

Un prompt multimodal no es un prompt de texto largo. Es una especificación de producción: qué se ve, cómo se ve, cómo se mueve, cómo suena, y qué no debe aparecer. Aquí sistematizamos las técnicas que separan un experimento bonito de una pieza publicable.

## A.1 · Anatomía de prompt visual: 7 capas

Todo prompt de imagen o video debe declarar explícitamente estas 7 capas. El orden importa porque los modelos pesan más lo que aparece primero.

Capa	Qué declarar
1. Sujeto	Quién/qué es el protagonista. Edad, género, etnia, expresión, vestuario, props.
2. Acción	Qué hace en el instante capturado. Verbo concreto, no abstracto.
3. Entorno	Dónde está. Lugar, hora del día, clima, elementos contextuales.
4. Encuadre	Plano (close-up, medium, wide), ángulo (low/high/eye), lente (35mm, 85mm, anamórfico), profundidad de campo.
5. Iluminación	Tipo de luz (natural golden hour, neón, rim light, hard studio), dirección, contraste.
6. Estilo	Referencia visual (cine de Wes Anderson, fotografía editorial Vogue, anime estudio Ghibli), paleta, textura.
7. Calidad técnica	Resolución, nivel de detalle, sharpness, grain, render engine cuando aplique.

### PROMPT Plantilla 7 capas

[Sujeto detallado] [acción concreta] in [entorno con hora/clima], [plano + ángulo + lente + DOF], [iluminación], [referencia de estilo + paleta], [calidad técnica].

### PROMPT Ejemplo aplicado

Mujer afrocolombiana de 32 años, sonrisa cálida, blusa lino crudo, tomando café · sentada en café-bar minimalista de Manizales al amanecer, ventanal grande · medium close-up, ángulo a la altura de los ojos, lente 50mm f/1.8, bokeh suave · luz natural golden hour entrando lateral · estilo editorial Kinfolk, paleta tierra y verde lima · 8K, grain 35mm sutil, foto-realismo.

## A.2 · Prompts negativos y exclusiones

Los modelos generativos producen artefactos predecibles. Bloquéalos siempre.

### **PROMPT** Negativo base que deberías reciclar

deformed hands, extra fingers, mutated anatomy, text artifacts, watermark, logo, low resolution, blurry, oversaturated, plastic skin, uncanny eyes, duplicate subject, cropped face.

## A.3 · Cinematografía y cámara

Variable	Vocabulario útil
Tipo de plano	Extreme wide, wide, medium, medium close-up, close-up, extreme close-up.
Ángulo	Low angle (poder), high angle (vulnerabilidad), eye-level (cercanía), Dutch tilt (tensión).
Movimiento (video)	Dolly in/out, push-in, pan L/R, tilt up/down, orbit, handheld, static.
Lente	24mm wide, 35mm doc, 50mm natural, 85mm retrato, 135mm compresión, anamórfico cinematográfico.
DOF	Shallow depth of field f/1.4 (bokeh fuerte), deep f/8 (todo en foco).

## A.4 · Style anchoring y consistencia de marca

La gran diferencia entre un experimento y una campaña es la coherencia visual entre piezas. Tres técnicas que sí funcionan:

- Seed fija: bloquea el seed para que micro-cambios en el prompt no rompan la composición.
- Imagen de referencia (style transfer): sube una pieza ancla y pide al modelo que respete paleta, textura y atmósfera.
- Character reference: usa la misma imagen del personaje como referencia en cada nueva escena (image-to-image como ancla).
- Naming de assets: marca\_v1\_personajeA\_escena01\_seed42.png — sin esto pierdes trazabilidad en 20 minutos.

### **TIP** Regla de oro de marca

Define 3 anclas no negociables (paleta, tipo de luz, encuadre dominante) y repítelas literalmente en cada prompt. Todo lo demás puede variar.

## A.5 · Iteración: refinar sin regenerar todo

Técnica	Cómo se aplica
Inpainting dirigido	Selecciona solo el área a corregir (mano deformada, fondo distractor) y reprompta solo esa región.
Keep X / Change Y	En image-to-image: "mantén composición, iluminación y personaje; cambia el fondo a [X]".
Upscaling con prompt	Al subir resolución, vuelve a especificar materiales y micro-textura para ganar detalle real, no solo píxeles.
Variations	Pide N variantes con seed distinto; quédate con la que tenga mejor composición y refínala.

## A.6 · Image-to-video: prompts de movimiento

En video con i2v no redescribas la imagen. Describe solo lo que se mueve y la intención de cámara.

### PROMPT Estructura i2v

Sujeto: [acción que ocurre – micro-movimiento concreto]. Cámara: [estática | dolly in lento | pan derecha suave]. Intensidad: [baja | media | alta]. Duración foco: [primeros 2s tensión, segundos 3-5 release].

### PROMPT Ejemplo i2v

Sujeto: el vapor del café sube en espiral, la mujer parpadea y sonríe levemente. Cámara: push-in muy lento de 5%. Intensidad: baja. Atmósfera contemplativa, sin cortes.

## A.7 · Prompts de voz y SFX (preludio Bloque B)

La voz también se prompteó. ElevenLabs v3 acepta Audio Tags en línea con el texto que dirigen interpretación. Se profundiza en el Bloque B; aquí queda el principio.

### PROMPT Ejemplo Audio Tags v3

[whispers] No te lo vas a creer... [pause] [excited] ;acaban de bajar 30% el precio!  
[laughs softly]

## A.8 · Cadenas multimodales: prompts puente

La magia está en cómo el output de una etapa alimenta el prompt de la siguiente. Plantilla de cadena:

**PROMPT Cadena copy → imagen → video → voz**

- 1) LLM (Manus/Claude): genera *headline + subhead + descripción visual de 3 líneas* a partir del brief.
- 2) Pasa la descripción visual al prompt 7-capas para imagen.
- 3) Convierte la imagen aprobada en video con prompt de movimiento.
- 4) Convierte el copy en guion de 12-18s y locútalo en ElevenLabs con Audio Tags acordes al tono del headline.

## A.9 · Versionado y QA de prompts

- Mantén una tabla viva con: prompt, modelo, seed, fecha, resultado (✓/×), por qué.
- Cuando algo funciona, etiquétalo como canon y réusalo como base.
- Compara siempre 3 variantes antes de quedarte con una. La primera nunca es la mejor.

## A.10 · Ejercicio de cierre del bloque

**EJERCICIO Ejercicio A · 20 min**

Toma tu brief real. Produce: (1) 1 headline + 1 subhead, (2) 1 prompt 7-capas para imagen hero, (3) 1 prompt i2v para esa misma imagen, (4) 1 guion de 15s con Audio Tags v3. Sin generar todavía — solo la especificación. Lo ejecutaremos en los Bloques B y C.

## BLOQUE B · 75 MIN

## ElevenLabs · ElevenCreative

ElevenCreative es la plataforma todo-en-uno de ElevenLabs: voz, música, SFX, imagen, video, edición y localización. Recorreremos cada módulo con una micro-tarea y al final tendrás la pista sonora completa de tu pieza.

**CLAVE** Qué es ElevenCreative en una frase

Una sola plataforma para generar, editar y localizar audio y video premium, con modelos nativos de ElevenLabs y 50+ modelos de imagen y video integrados.

### B.1 · Text to Speech v3 (Audio Tags)

v3 es el modelo de TTS más expresivo disponible: ritmo humano, respiración, emoción e inflexión en 70+ idiomas. Su superpoder son los Audio Tags en modo Expressive.

#### Audio Tags más usados

Categoría	Tags útiles
Emoción	[laughs], [laughs softly], [sighs], [excited], [sad], [angry], [whispers], [shouting]
Ritmo	[pause], [long pause], [slow], [fast]
Énfasis	[emphasis], en mayúsculas para gritar, comillas para citar
Sonido	[clears throat], [breathes in], [chuckles]

#### Controles que sí mueven el resultado

Parámetro	Cuándo subirlo / bajarlo
Stability	Bajo = más expresivo y variable. Alto = más consistente y plano.
Similarity	Qué tan fiel a la voz original. Alto para clones, medio para voces de librería.
Style	Cuánto exagerar el estilo de la voz. Sutil para narración, alto para personajes.
Speed	Velocidad de habla. 0.9-1.1 es zona segura.

**EJERCICIO Ejercicio B.1**

Locuta tu guion del Bloque A en español neutro y en una voz expresiva. Genera 3 versiones cambiando solo Stability (0.3 / 0.5 / 0.7). Quédate con la mejor.

## B.2 · Voice Library, Voice Design y Voice Changer

- Voice Library: 10.000+ voces. Filtra por idioma, género, edad, caso de uso. Previsualiza antes de generar.
- Voice Design: crea una voz nueva desde un prompt textual o sliders (edad, tono, acento, personalidad). Ideal cuando no encuentras la voz exacta.
- Voice Changer (Speech to Speech): graba tu interpretación con tu voz y conviértela a otra preservando emoción y ritmo. Útil cuando el TTS no te da el matiz que necesitas.

**PROMPT Voice Design — prompt ejemplo**

Voz masculina, 40 años, acento colombiano paisa neutro, cálida, con ligera sonrisa, tono cercano de podcast nocturno, ritmo pausado.

## B.3 · Voice Cloning

Modo	Cuándo usarlo
Instant Cloning	60 segundos de audio. Para uso personal y prototipos rápidos.
Professional Cloning	30+ minutos de audio limpio. Calidad de producción, multilingüe, ideal para marca propia o talento contratado.

**ATENCIÓN Consentimiento**

Solo clona voces propias o con autorización explícita y documentada del titular. Es requisito legal y de los Términos de ElevenLabs.

## B.4 · Sound Effects

Genera SFX desde texto: ambientes, foley, transiciones, impactos. La descripción ideal incluye fuente del sonido + entorno + duración aproximada + intensidad.

**PROMPT SFX — ejemplos**

- Crujido suave de hojas secas pisadas, bosque húmedo al amanecer, 2 segundos, intensidad media.
- Whoosh cinematográfico grave para transición, 1.5 segundos, con cola de reverberación.
- Ambiente café-bar lleno, conversaciones de fondo apenas audibles, máquina de espresso ocasional, loop de 30 segundos.

## B.5 · Music

Música original generativa, con o sin voz, en múltiples géneros e idiomas. Ajustable en duración y looping. Royalty-free para uso estándar; campañas de marketing, publicidad, cine, TV y games requieren licencia comercial adicional.

### PROMPT Music — prompt ejemplo

Indie folk acústico, guitarra arpegiada cálida, percusión suave de cajón, sin voz, 60 BPM, mood esperanzador y nostálgico, build-up a partir de los 0:20, duración 0:45, fade-out.

## B.6 · Dubbing y Scribe (STT)

- Dubbing: traduce y dobla audio o video preservando tono, emoción y timing del hablante. Clona automáticamente la voz para mantener coherencia entre idiomas. 100+ idiomas con localización (no traducción literal).
- Scribe (Speech to Text): transcribe con timestamps, diarización de hablantes y etiquetado de eventos sonoros. 99+ idiomas. Útil para subtítulos y contenido derivado (blog, redes).
- Voice Isolator: limpia voz quitando ruido, música o SFX de fondo.

## B.7 · Studio

Workspace de producción con timeline editor: pistas dedicadas para video, narración, música, SFX y captions. Importas footage, generas narración, montas, agregas captions y exportas como audio o video terminado.

### Sub-flujos clave en Studio

Sub-flujo	Para qué
Audiobook	Sube un documento, elige voz (librería o clonada) y produce audiolibro completo. Publicación directa en Spotify y retailers vía ElevenReader.
GenFM	Genera podcast con IA a partir de cualquier material fuente.
AI Soundtrack	Genera banda sonora que matchea mood y pacing de tu video automáticamente.
Actor Mode	Dirige la interpretación usando tu voz como referencia de cómo quieres que se diga la línea.

## B.8 · Flows

Workspace nodal en lienzo infinito. Conectas nodos de imagen, video, voz, música, lip-sync y SFX en secuencia, y ejecutas todo el pipeline en un solo run. Soporta iteración no destructiva: rehaces un paso sin regenerar el resto. Ideal para A/B testing y rollouts multi-idioma a escala.

**TIP Cuándo Flows vs Studio**

Studio = edición fina de una pieza. Flows = pipeline reutilizable que produce muchas variantes (cambia hook, producto, idioma, avatar y batch genera).

## B.9 · Image & Video y Productions

- Image & Video: genera imágenes desde texto y videos desde texto o imagen usando 50+ modelos líderes, todo desde dentro de ElevenCreative.
- Productions: ordena transcripciones, subtítulos, doblajes y audiolibros editados por humanos (lingüistas vetados de ElevenLabs) cuando necesitas calidad de agencia a fracción del costo.

## B.10 · Ejercicio integrador del bloque

**EJERCICIO Ejercicio B · 25 min**

1) Locuta tu guion en TTS v3 con Audio Tags. 2) Genera 1 SFX que apoye el guion. 3) Genera una pista de música de 30s que matchee el mood. 4) Mezcla las 3 pistas en Studio y exporta MP3 final. Bonus: dobla el resultado al inglés con Dubbing.

## BLOQUE C · 60 MIN

# FloraFauna.ai

FloraFauna (flora.ai) es un lienzo creativo nodal donde encadenas modelos de imagen y video de última generación. Es donde tus prompts del Bloque A se vuelven píxeles y movimiento.

## C.1 · Canvas y nodos

- Lienzo infinito con nodos de generación, edición y conexión.
- Cada nodo guarda su prompt, modelo, seed y referencias — trazabilidad total.
- Ramificas variantes sin perder la versión original (no destructivo).
- Comparas resultados lado a lado en el mismo canvas.

## C.2 · Modelos de imagen integrados

Modelo	Para qué brilla
Flux (Pro / Dev / Schnell)	Foto-realismo y consistencia general. Pro = máxima calidad.
Ideogram	El mejor para texto legible dentro de la imagen (logos, posters, packaging).
Recraft	Vector e ilustración con control de estilo y paleta consistente.
SDXL y otros	Stock generalista; útil para variantes rápidas.

### TIP Regla práctica de selección

Foto de marca real → Flux Pro. Pieza con texto/typo → Ideogram. Ilustración o icon set → Recraft. Concept art rápido → SDXL.

## C.3 · Image-to-image, inpainting y upscaling

- Image-to-image: usa una imagen como base y reprompta para variar estilo, escena o personaje conservando composición.
- Inpainting: pinta una máscara sobre la zona a cambiar y reprompta solo esa región (mano, fondo, detalle de producto).
- Upscaling: sube resolución a 2x/4x reespecificando materiales para ganar detalle real.

### EJERCICIO Ejercicio C.1

Genera la imagen hero de tu campaña aplicando tu prompt 7-capas. Saca 4 variantes, elige una, refínala con inpainting (corrige un detalle) y upscaléala a 2x.

## C.4 · Mezcla de referencias y consistencia de personaje

Sube 1-3 imágenes de referencia y asigna pesos: una para estilo visual, otra para personaje, otra para composición. Así garantizas que tu modelo/producto luzca igual en cada pieza de la campaña.

## C.5 · Image-to-video (Kling, Hailuo y más)

Modelo de video	Fortaleza
Kling	Movimiento natural, físicas creíbles, buen lip-sync básico, 5-10s.
Hailuo (MiniMax)	Cámara cinematográfica, planos largos coherentes.
Runway / Luma	Estilizados, transiciones limpias para social.

### PROMPT Plantilla i2v Flora

Imagen base: [tu hero]. Movimiento: [micro-acción del sujeto]. Cámara: [estática / push-in lento / pan]. Intensidad: baja. Duración: 5s.

## C.6 · Flujos visuales encadenados en el canvas

La práctica avanzada: en un mismo canvas, encadenas modelo de imagen → inpaint → upscale → image-to-video → variante en otro modelo de video. Cada nodo es un experimento documentado.

### EJERCICIO Ejercicio C · 25 min · entregable de la jornada

En tu canvas: (1) Hero estática refinada (imagen + inpaint + upscale). (2) 2 piezas estáticas adicionales para carrusel manteniendo paleta y personaje. (3) 1 animación de 5s sobre la hero (i2v). Etiqueta cada nodo. Exporta los 3 archivos.

CIERRE · 15 MIN

## Flujo end-to-end y entrega

Acabas de producir, en 4 horas, los activos de una mini campaña multimodal. Recapitulemos el flujo que dominaste y cómo replicarlo cualquier lunes.

### Flujo canon

Etapa	Acción
1. Brief	Cliente, producto, objetivo, audiencia, tono, KPI.
2. Copy	LLM (Manus/Claude/ChatGPT) → headline, subhead, guion 15s, descripción visual.
3. Imagen	FloraFauna · Flux/Ideogram con prompt 7-capas → variantes → inpaint → upscale.
4. Video	Flora i2v (Kling/Hailuo) con prompt de solo movimiento.
5. Voz	ElevenLabs TTS v3 con Audio Tags acordes al tono del copy.
6. SFX + Música	ElevenLabs SFX + Music alineados a mood y pacing.
7. Mezcla	Studio → timeline → export.
8. Localización	Dubbing 100+ idiomas si la campaña lo pide.

#### CHECKLIST Checklist de entrega del día

Brief documentado · Headline + subhead + guion · Hero estática refinada · 2 piezas adicionales · 1 video de 5s · 1 locución con Audio Tags · 1 SFX · 1 pista de música · MP4 final mezclado · Tabla de prompts versionada

### Para seguir practicando

Herramienta	Acceso
ElevenLabs (acceso Confa)	<a href="https://try.elevenlabs.io/confa">https://try.elevenlabs.io/confa</a>
FloraFauna.ai	<a href="https://flora.ai/">https://flora.ai/</a>
Manus (invitación)	<a href="https://manus.im/invitation/KKHKSDN9TIY5TXU">https://manus.im/invitation/KKHKSDN9TIY5TXU</a>

**CLAVE La habilidad real que te llevas**

No son las herramientas — cambian cada 6 meses. Es la capacidad de especificar producción con precisión: 7 capas, prompts de movimiento, Audio Tags, cadenas multimodales y versionado. Eso se queda contigo en cualquier stack.