

Coloque o que você quiser dentro de um vídeo **que já está gravado**

Um bicho, um produto, uma pessoa passando ao fundo. Entra na cena como se estivesse lá na hora da filmagem, e o seu rosto, a sua voz, a sua fala, o lip sync, os gestos, o cenário e o movimento da câmera **continuam exatamente iguais**.

Não é gerar um vídeo novo com o elemento dentro. É dizer ao modelo que o vídeo é real, que ele fica como está, e que uma única coisa vai ser acrescentada. Essa diferença é o método inteiro.

4 a 30s

duração aceita pelo modelo

9 créditos

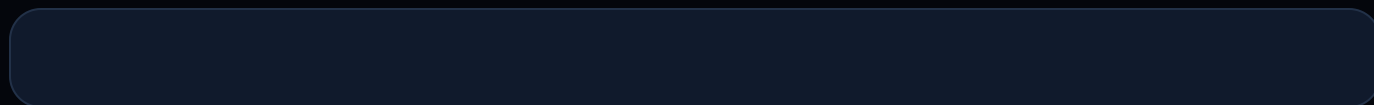
por segundo, em 1080p

1 geração

quando o prompt está certo

O antes e o depois

Repare que a fala, o rosto, os gestos e o enquadramento continuam idênticos. A única coisa nova é o elemento.



A foto que trava a aparência

São sempre duas entradas: o seu vídeo e uma foto do que vai entrar. É a foto que faz o desenho sair certo **já na primeira tentativa**.

Sem a foto, o modelo inventa o elemento do zero a cada geração, e você nunca consegue repetir nem corrigir. Com ela, ele copia cor, textura, proporção e desenho.

O que faz a foto funcionar

- + O elemento sozinho: tudo que estiver na foto pode entrar no vídeo
- + Corpo inteiro, porque close da cabeça devolve um corpo inventado
- + De perfil ou de três quartos: mostra silhueta e volume ao mesmo tempo
- + Luz parecida com a da cena de destino

O que estraga

- ✗ Mão, dedo ou braço segurando: vira parte do elemento
- ✗ Outra pessoa no quadro: pode aparecer no vídeo
- ✗ Cenário forte atrás, como parede de tijolo: o modelo puxa junto
- ✗ Marca d'água, texto ou logo: são reproduzidos

Passo a passo, do arquivo bruto ao vídeo pronto

São 11 passos e o seu progresso fica salvo neste navegador. Os que mais mudam o resultado são **o 3, o 5 e o 7**.

1 Confira o que precisa estar instalado

São três coisas, e você provavelmente já tem duas. O FFmpeg corta, transcodifica e recola áudio. O Python roda o analisador. O acesso ao modelo vem pelo Higgsfield, que é onde a edição acontece de verdade.

COPIA E COLA

```
ffmpeg -version
ffprobe -version
python --version

:: se faltar o ffmpeg:
winget install Gyan.FFmpeg

:: conectar o Higgsfield ao Claude Code:
claude mcp add higgsfield https://mcp.higgsfield.ai/mcp
```

Atalho: No Windows use `python`, nunca `python3`. O `python3` aqui é um atalho da Microsoft Store que não executa nada, e esse é o primeiro lugar onde todo mundo trava.

2

Escolha um vídeo que respeite os três limites

Entre 4 e 30 segundos, uma tomada contínua, e o lugar onde o elemento vai entrar precisa aparecer em algum momento. O terceiro limite é o mais ignorado e o que mais estraga resultado: inserir numa área que o vídeo nunca mostrou obriga o modelo a inventar cenário.

1. Duração entre 4 e 30 segundos. Maior que isso, fatie em blocos
2. Uma tomada só, sem corte de edição no meio
3. O lugar da inserção aparece em quadro em algum trecho
4. Tem um objeto de tamanho conhecido por perto, que vira a âncora de escala

Atalho: Se a câmera vira no meio do vídeo, isso não impede a inserção. Mas o elemento só existe enquanto a área dele está em quadro, e o resto precisa ser blindado no prompt.

3 Analise a cena antes de escrever qualquer coisa

Este passo separa quem acerta de primeira de quem queima crédito. O script extrai os quadros, lê a duração e monta o plano. Depois você olha os quadros, um por um. Não deduz a cena pela transcrição.

1. Onde o elemento cabe: a superfície, a profundidade, o que já ocupa aquele espaço
2. Objeto âncora de escala: algo de tamanho conhecido perto da inserção
3. O que fica na frente: vidro, folha, tronco, borda de mesa, ombro
4. A luz: direção, dureza, temperatura
5. Tudo que precisa ser preservado: roupa, acessórios, cenário

COPIA E COLA

```
python "%USERPROFILE%\claude\skills\videos-inserir-elementos\materiais\scripts\analisar_cena.py"  
"CAMINHO\DO\VIDEO.mp4"
```

Atalho: Foi olhando quadro a quadro que apareceu, num caso real, uma virada de câmera aos 9,6 segundos que a transcrição não mostrava. Sem ter visto aquilo, a geração inteira teria ido para o lixo.

4 Cruze a fala com a imagem e ache a janela do elemento

Responda duas perguntas. De que segundo a que segundo a área da inserção está em quadro? Fora dessa janela o elemento não existe. E a fala aponta para o elemento em algum momento? Se você diz olha isso, é nesse segundo que ele precisa estar bem visível.

Atalho: Esse casamento entre fala e elemento não custa nada e é o que faz a inserção parecer intencional, como se o vídeo tivesse sido pensado para aquilo desde o começo.

5 Prepare a foto de referência

A foto vai junto com o vídeo e é ela que trava a aparência do elemento. Sem foto, o modelo inventa o elemento do zero a cada geração e você nunca consegue repetir nem corrigir. É o passo que mais muda o resultado e o que mais gente pula.

1. Elemento sozinho, sem mão segurando e sem outra pessoa no quadro
2. Corpo inteiro, porque close da cabeça devolve um corpo inventado
3. Fundo simples: fundo forte é copiado junto
4. Sem marca d'água, texto ou logo, porque são reproduzidos

COPIA E COLA

```
:: 1. varredura para escolher o quadro
ffmpeg -v error -y -i referencia.mp4 -vf "fps=1/2,scale=540:-1" frames/ref_%02d.jpg

:: 2. extraia o melhor em resolucao cheia
ffmpeg -v error -y -ss 17 -i referencia.mp4 -frames:v 1 ref_full.png

:: 3. recorte so o elemento e suba a resolucao
ffmpeg -v error -y -i ref_full.png -vf "crop=500:230:170:590,scale=1000:460" elemento-ref.png
```

Atalho: Olhe o recorte antes de subir. Tudo que estiver na foto pode entrar no vídeo. Dois minutos recortando economizam uma geração inteira.

6 Prepare o vídeo base

Vídeo de celular costuma vir em 4K, que é pesado e desnecessário: a saída sai em 1080p de qualquer jeito. Transcodifique antes de subir e guarde o original, porque é dele que sai o áudio no final.

COPIA E COLA

```
ffmpeg -v error -y -i base-original.mov ^  
-vf "scale=1080:1920" -c:v libx264 -crf 20 -preset slow -pix_fmt yuv420p ^  
-c:a aac -b:a 192k base-1080.mp4  
  
:: para 16:9 use scale=1920:1080
```

Atalho: Num caso medido isso levou o arquivo de 33 MB para 14 MB, sem perda visível na saída.

7 Escreva o prompt na proporção 80/20

Esta é a ideia central do método inteiro: 80 por cento do prompt protege o que já está no vídeo, 20 por cento descreve o que entra. Se o seu ficou ao contrário, o resultado deforma. Você não está gerando um vídeo novo com o elemento. Você está dizendo que o vídeo é real, que ele fica exatamente como está, e que uma única coisa vai ser acrescentada.

1. Escreva em inglês: o modelo obedece melhor. A fala do vídeo continua no idioma original
2. Timecodes explícitos que somam a duração real do vídeo
3. Ancore a escala num objeto que já aparece na cena, nunca a palavra small
4. Declare a profundidade e repita: BEHIND the glass, never in front of it
5. Nomeie o que oclui: folhas, tronco, borda, moldura, ombro
6. Blinde todo trecho onde o elemento não aparece, com uma linha própria
7. Nunca nomeie equipamento de câmera: o modelo desenha o equipamento dentro da cena

Atalho: O template completo com os 8 blocos está mais abaixo nesta página, pronto para copiar.

8 Suba os arquivos e faça o preflight de custo

São duas entradas, sempre as duas: o vídeo e a foto. Antes de gerar, rode a mesma chamada com `get_cost`, que não submete job nenhum e devolve o preço. Guarde os dois identificadores: para corrigir depois você não reenvia arquivo.

COPIA E COLA

```
generate_video({ params: {  
  model: "seedance_2_5",  
  mode: "video_edit",  
  resolution: "1080p",  
  use_unlim: false,  
  medias: [  
    { role: "video_references", value: "<media_id do video>" },  
    { role: "image_references", value: "<media_id da foto>" }  
  ],  
  get_cost: true  
}})
```

Atalho: Rascunhe em 720p enquanto acerta a direção e vá para 1080p só na versão final: economiza cerca de 28 por cento.

9 Recuse o preset e gere

A plataforma quase sempre devolve uma sugestão de preset antes de aceitar o seu texto, com um nome aleatório que nada tem a ver com o pedido. Aceitar descarta a sua direção inteira. Recuse e repita a chamada com o prompt literal.

Atalho: Um vídeo de 12 segundos leva de 5 a 12 minutos para renderizar. Acompanhe com a chamada enxuta de status, não com a que devolve o prompt inteiro.

10 Recole o áudio original

O resultado volta com a fala recriada pelo modelo. Não é a sua voz, e às vezes a última palavra sai trocada. Isso não se resolve escrevendo melhor o prompt. O vídeo também volta um pouco mais curto e com taxa de quadros diferente, então o script estica para a duração original e o lip sync bate de novo.

COPIA E COLA

```
python "%USERPROFILE%\claude\skills\videos-inserir-elementos\materiais\scripts\recolar_audio.py"
gerado.mp4 base-original.mp4 final.mp4
```

Atalho: Medido: um base de 12,667s a 30fps voltou com 12,384s a 24fps. O ajuste ficou em 1,022x, imperceptível, e devolveu a voz real com as palavras certas.

11 Confira olhando, não pelo ffprobe

Extraia quadros nos pontos-chave do arco e olhe cada um. Quem valida identidade e lip sync é o olho humano, então não dê por certo sem ter olhado.

1. A escala bate com os objetos ao redor em todos os quadros
2. Está no plano de profundidade certo, nunca na frente do que deveria estar atrás
3. Toca a superfície e tem sombra de contato, não flutua
4. O rosto está idêntico do começo ao fim, sem variação
5. Nos trechos blindados o elemento realmente não aparece

COPIA E COLA

```
for %t in (1.0 2.5 4.0 5.5 7.0 9.2 11.5) do ffmpeg -v error -y -ss %t -i final.mp4 -frames:v 1 -
vf scale=540:-1 check\g_%t.jpg
```

Atalho: Guarde o prompt usado e os dois identificadores num arquivo ao lado do vídeo. Sem isso, corrigir depois vira começar do zero.

As quatro âncoras

Se você só puder acertar quatro coisas no prompt, são estas. Cada uma corresponde a **um defeito clássico**.

ÂNCORA	COMO SE ESCRIVE	DEFEITO QUE EVITA
Escala	<code>roughly two to three times the height of the small red monk figurine</code>	elemento gigante ou minúsculo
Profundidade	<code>BEHIND the glass panel at all times, never in front of it</code>	elemento pulando para o primeiro plano
Fotografia	<code>slightly lower in contrast and desaturated, matching how the plant already looks through that same glass</code>	elemento nítido demais, parecendo colado
Oclusão	<code>occluded by the leaves and the trunk when it passes behind those</code>	elemento flutuando sobre o cenário

Atalho: a escala vai sempre amarrada num objeto que já aparece no vídeo. A palavra "small" não significa nada para o modelo. A estatueta significa.

O arco de comportamento

Elemento parado convence menos que elemento com comportamento. O que faz o olho aceitar não é o desenho, é **a física dele e a relação com a cena**.

1. Presença (30 a 40%)

Já está lá quando o vídeo começa, ocupado com alguma coisa própria. Não entra em cena: já estava.

2. Deslocamento (cerca de 20%)

Atravessa o quadro e sai. É o que prova que ele ocupa espaço de verdade.

3. Ausência (1 a 2 segundos)

A cena volta a ser exatamente a original. É a fase mais importante e a que todo mundo pula: sumir e voltar prova que aquilo é um ser no espaço, não um adesivo colado no fundo.

4. Retorno e pouso (25 a 30%)

Volta, chega perto e para, olhando para a câmera. Se a fala aponta para o elemento, o pouso vai nesse segundo exato.

Por tipo: bicho pequeno se move como ave, com passos curtos e paradas o tempo todo, porque bicho não desliza. Animal grande precisa de peso, com o corpo balançando a cada passo. Produto parado não se move sozinho: o único movimento nele é a luz mudando conforme a câmera respira. Pessoa ao fundo é o caso mais arriscado, e a regra é longe, de costas ou de perfil, atravessando.

O template do prompt

São 8 blocos e cada um existe por um motivo. Escreva em inglês, substitua o que está entre sinais de menor e maior, e apague os comentários.

TEMPLATE COMPLETO

VIDEO EDIT - ADD <O ELEMENTO>. CHANGE NOTHING ELSE.

This is a real <DURAÇÃO> second <vertical/horizontal> video, filmed handheld by <quem> in <onde>. Keep the entire video exactly as it is and add only one thing: <o elemento, em uma linha, com o lugar>. The attached reference image shows exactly what <o elemento> must look like.

CRITICAL - EVERYTHING BELOW MUST STAY 100% UNCHANGED

- <A pessoa>'s face, skin, beard, hair, eyes, expressions and every micro-movement of the face. Identity must not drift for a single frame.
- The voice, the speech in <idioma>, the exact words, the timing, the lip sync, the ambient sound. Do not re-record, re-time or replace the speech.
- Every gesture of the hands, on exactly the same frames as the original.
- <Roupa, acessórios, cada um pelo nome - camiseta preta com logo redondo no peito, fone branco na orelha>.
- The framing, the lens, the handheld camera and all of its original movement. No new camera angles, no cuts, no transitions, no zooms, no reframing, no slow motion, no speed change.
- The color, exposure, white balance, the <tipo de luz - flat overcast daylight> and the grain of the original footage.
- The entire set: <liste item por item - parede, planta, pedras, móveis, objetos pequenos, o que se vê pela janela>.
- <Se houver trecho com outra cena ou outro enquadramento, descreva e diga que fica intocado>.

THE ONLY ADDITION - <O ELEMENTO>

<Quantos> <o elemento> <verbo de existência> <onde exatamente, com referência a coisas que já estão na cena>. <Eles são/Ele é> real, filmado pela mesma câmera.

How <it looks / they look>, matching the reference image: <descrição física completa - tamanho em cm, forma, cor, textura, partes, detalhes de desenho>. Realistic <categoria>, not a toy, not a cartoon, not a plastic figure, not CGI.

How <it moves / they move>: <física do movimento - velocidade, peso, ritmo, o que a articulação faz, o que o corpo faz pra se equilibrar>.

THEIR ACTION, WITH TIMECODES

(cada bloco diz o que o elemento faz E confirma o que o vídeo original está fazendo naquele intervalo. Os timecodes têm que somar a duração real.)

00:00.0-00:0X.X - <presença: já está lá, ocupado com algo>

00:0X.X-00:0Y.Y - <deslocamento: atravessa e sai de quadro>

00:0Y.Y-00:0Z.Z - <ausência: a cena volta a ser exatamente a original, nada além do que já existe>

00:0Z.Z-00:0W.W - <retorno e pouso: volta, chega perto, para e olha>

00:0W.W-<FIM> - <BLINDAGEM: o elemento está fora de quadro daqui em diante. Nothing is added to this last part.>

PHYSICS AND INTEGRATION

- <Plano de profundidade, repetido>: it is BEHIND <o que> at all times, never in front of it, never <onde não pode estar>.
- <Casamento fotográfico>: because it is seen through <vidro/distância>, it is slightly lower in contrast and slightly desaturated compared to the foreground, matching how <algo que já está lá> already looks.
- <Contato físico>: its feet actually touch and displace <a superfície>. It casts soft contact

shadows under the same <tipo de luz> as the rest of the scene.

- <Oclusão nomeada>: it is correctly occluded by <folhas, tronco, borda, moldura, ombro> when it passes behind those.
- <Escala ancorada>: its scale is consistent with <objetos da cena> - roughly <N> times the height of <objeto âncora que já aparece no vídeo>.
- Motion blur, grain, focus and depth of field on it match the original handheld footage. It must look filmed, not composited.

NEGATIVES

No change to the face. No identity drift. No change to the voice, the words or the timing. No new speech, no narration, no sound effects. No <elemento> <onde ele não pode estar>. No <elemento> at all in <o trecho blindado>. No more than <N>, never <N+1>. No <versão errada do elemento - adult, giant, toy, plastic, cartoon, CGI look>. No cuts, no transitions, no morphs, no new camera angles, no zoom, no reframing, no slow motion. No text, no captions, no logos, no watermarks. No relighting, no color grading change, no stylisation. No camera equipment visible anywhere.

FINAL FEEL

It must look like <quem> really filmed this on <onde> with <o elemento> <fazendo o quê>, and that nothing at all was done to the footage.

O prompt literal do seu vídeo

Depois da sua primeira geração, cole aqui o prompt exato que funcionou, sem limpar nada. Serve de calibragem do nível de detalhe para quem está lendo.

Quando sai errado, e por quê

Cada defeito tem uma causa específica. Corrija só o bloco responsável, nunca o prompt inteiro, senão você perde a referência do que causou o quê.

SINTOMA	CORREÇÃO
Sai grande ou pequeno demais	Ancore num objeto da cena: duas a três vezes a altura da estatueta, nunca "small"
Aparece na frente do que devia estar atrás	Declare e repita: BEHIND the glass at all times, never in front of it
Parece adesivo colado no fundo	Nomeie o que passa na frente dele e inclua 1 a 2s em que ele sai de quadro
Vaza para o trecho final do vídeo	Uma linha de blindagem por intervalo vazio: nothing is added to this part
Sai um brinquedo ou boneco de plástico	Negativas explícitas: realistic living animal, not a toy, not a cartoon
Entra um terceiro elemento	Trave a quantidade: no more than two, never a third
O rosto muda sutilmente	Reforce a seção CRITICAL, com roupa e acessórios pelo nome
Muda a cor ou o contraste do vídeo todo	No relighting, no color grading change, no stylisation
Aparece um braço robô, grua ou tripé	Você nomeou equipamento no prompt. Descreva o ponto de vista, nunca a máquina
Saiu um preset genérico no lugar da sua direção	A sugestão de preset foi aceita. Recuse e gere de novo com o prompt literal
A última palavra da frase saiu trocada	O áudio foi recriado. Não tente consertar por prompt: recole o áudio do original

Atalho: se duas tentativas no mesmo ponto não resolverem, o problema costuma ser o lugar da inserção, não a redação. Normalmente é uma posição que exige inventar cenário

que nunca apareceu. Mova o elemento para uma área que já está em quadro.

Quanto custa, de verdade

Custo linear na duração, e a resolução conta. Estes números foram **medidos, não estimados**.

9 créditos/s

em 1080p. Um vídeo de 12,7s saiu por 114 créditos

6,5 créditos/s

em 720p. O mesmo vídeo sai cerca de 28% mais barato

Igual

cada correção custa o mesmo que a primeira geração

- + Rascunhe a direção em 720p e vá para 1080p só quando o arco estiver certo
- + Rode sempre o preflight de custo antes de gerar: é a mesma chamada e não submete nada
- + Como cada tentativa custa igual, vale gastar dois minutos a mais no prompt
- + Guarde os dois identificadores: para corrigir depois você não reenvia arquivo nenhum

A skill pronta, para instalar no Claude Code

Tudo isso empacotado, com os scripts **já corrigidos para Windows**.

Vai dentro o analisador de cena, o script que recola o áudio, o template de prompt, a biblioteca de comportamentos, o guia da foto de referência e o checklist de qualidade.

Como instalar: descompacte em %USERPROFILE%\Claude\skills\ e reinicie o Claude Code. Depois é só pedir em português o que você quer inserir, e a skill escreve o prompt inteiro sozinha.

O que dá certo e o que não dá

Funciona bem

- + Vídeo de 4 a 30 segundos, uma tomada contínua
- + Elemento num lugar que já aparece em quadro
- + Bicho, produto, objeto, algo se movendo ao fundo
- + Cena com objeto de tamanho conhecido por perto

Não tente

- ✗ Inserir numa área que o vídeo nunca mostrou: o modelo inventa cenário
- ✗ Vídeo maior que 30s sem fatiar em blocos
- ✗ Confiar no áudio que volta: ele é sempre recriado
- ✗ Rosto humano novo em primeiro plano, olhando para a câmera

⚡ Próximo passo

Isso aqui é um pedaço do que dá para automatizar

Inserir elemento em vídeo é uma tarefa. O Operação Comando monta a estrutura inteira, com agentes de IA trabalhando na sua operação todos os dias.

Conhecer o Operação Comando

[Operação Comando](#) | [Privacidade](#) | [Termos](#)