MiniMax H3 導入手順書

株式会社スターストリーム・スタジオ　v1.0（2026年10月）

動画AI「Hailuo 3.0」と同じモデルとして公開された MiniMax H3 を、自分のパソコン（ComfyUI）で動かすための手順書です。必要なパソコンの条件、入れ方、最初の1本の作り方、日本語のセリフを話させる書き方、広告に使うときのライセンスの確認ポイントまでを順番に載せています。

もくじ

1. MiniMax H3 でできること
2. 必要なパソコンの条件
3. 広告に使う前に、ライセンスを確かめる
4. ComfyUI を用意する
5. モデルを入れる（5つのファイル）
6. テンプレートを開く
7. 最初の1本を作る
8. 日本語のセリフを話させる
9. 止まらずに流すための設定
10. つまずきやすいところ
11. かかる時間とお金の目安

---

1. MiniMax H3 でできること

- **画像1枚から、4〜15秒の動画を作ります。**縦長（9:16）・横長（16:9）・正方形などに対応しています
- **映像と一緒に、声・効果音・音楽も出ます。**日本語を含む11言語のセリフを安定して話せます
- **短い辺が768ピクセル・1秒24コマ**で出ます
- **自分のパソコンで動くので、API 代はかかりません。作れる本数に上限もありません**

（出典: MiniMax H3 のモデルページ https://huggingface.co/MiniMaxAI/MiniMax-H3 ・2026年10月確認）

2. 必要なパソコンの条件

公式に書かれていること

MiniMax H3 のモデルページと ComfyUI の公式の解説には、**必要な GPU のメモリやメインメモリの数字は書かれていません**（2026年10月確認）。書かれているのは次の点です。

- モデルの大きさは **33B（330億パラメータ）**
- ComfyUI の版は **0.30.0 以上**（画像から動画・文字から動画のテンプレート）
- **int8 と fp8 の軽い版**が用意されていて、bf16 版は「より多くの GPU メモリが要る」
- 速くする工夫（sparse attention など）は **NVIDIA の CUDA** が前提

実際に動いた構成（私の環境・実測）

部品	中身
OS	Windows 11
GPU	**NVIDIA GeForce RTX 4070 Ti SUPER（GPU のメモリ 16GB）**
メインメモリ	**32GB**
保存場所	SSD。モデルだけで**約42GB**（ComfyUI 本体と環境を入れて60GB ほど空けておくと安心）
ComfyUI	v0.40.0（2026年10月13日の版）
PyTorch	2.14.1＋cu130（CUDA 13）

目安

- **GPU：NVIDIA で、GPU のメモリが 16GB 以上**
- **メインメモリ：32GB 以上。**32GB では、生成している間はほぼ使い切ります。ブラウザなどほかのアプリを閉じて流してください。64GB あると余裕が出ます
- **SSD に 60GB 以上の空き**（モデルの読み込みが速いため、HDD より SSD）
- Mac・AMD の GPU・GPU のメモリが 12GB 以下の環境は、私は試していません

---

3. 広告に使う前に、ライセンスを確かめる

MiniMax H3 は「MiniMax H3 Community License」で公開されています。広告や商品に使う前に、原文（モデルページの LICENSE）を読んでください。私が原文で確かめた要点は次のとおりです。

- **日本は利用できる地域に入っています**（対象外は EU・英国・韓国・米国の4つ）
- **公開する動画は、機械で作ったものだと分かるようにする必要があります。**広告なら「この動画はAIで作成しています」のような表示を入れます
- 商用の製品・サービスに組み込むときは「MiniMax H3」を目立つ所に表示します
- **年商2,000万ドルを超える会社は、別の契約が要ります**
- 他人になりすます使い方はできません

※ 規約は変わることがあります。使う時点の原文を必ず確かめてください。

---

4. ComfyUI を用意する

1. ComfyUI を入れます。すでに入っている場合は、**0.30.0 以上**に更新します（私は v0.40.0 に上げました）
2. 起動したときに「速い計算には新しい CUDA が要る」という表示が出たら、ComfyUI の環境の PyTorch を新しい版（cu130）に入れ替えます。入れ替えたあと起動し直すと、CUDA の部品が有効になります
3. 更新する前に、今の状態に戻せる印を付けておくと安心です（git で入れている場合は、今のコミットにタグを付ける）

⚠ ほかの動画モデル（Wan など）を入れている場合、ComfyUI を更新すると、その環境も一緒に変わります。使わないモデルは先に片づけておくと、保存場所も空きます。

---

5. モデルを入れる（5つのファイル）

公式の置き場所は Hugging Face の **Comfy-Org/MiniMax-H3** です。いちばん軽い組み合わせで、次の5つを入れます（合わせて約42GB）。

置く場所（ComfyUI の models の中）	ファイル	大きさ	役割
diffusion_models	`minimax_h3_fl2va_pruned_int8_convrot.safetensors`	21.0GB	本体（画像から動画・軽い int8 版）
text_encoders	`qwen3vl_32b_minimax_h3_nvfp4_awq.safetensors`	15.7GB	頼み方の文章を読む役
vae	`minimax_h3_video_vae_int8_convrot.safetensors`	2.8GB	映像の部品
vae	`minimax_h3_audio_vae_fp32.safetensors`	0.6GB	音の部品
loras	`minimax_h3_fl2v_turbo_8step_v1.0_comfyui_bf16.safetensors`	2.0GB	8ステップで速く作る部品

- ダウンロードは1本ずつより、いくつか同時に落としたほうが早く終わります
- ファイル名が1文字でも違うと、テンプレートが見つけられません。名前は変えずに置きます

---

6. テンプレートを開く

1. ComfyUI を起動し、**テンプレートの一覧** → **Video** → **MiniMax H3** を開きます
2. 画像から動画を作るときは **Image to Video**（`video_minimax_h3_i2v`）を選びます
3. 🔴 **名前に「api」が付いたテンプレート**（`api_minimax_h3_…`）**は使いません。**クラウドの有料の経路で、使った分だけ料金がかかります
4. テンプレートが一覧に無いときは、ComfyUI の版が古いか、起動のときに部品の読み込みに失敗しています

---

7. 最初の1本を作る

最初は、小さく短い動画で「最後まで通るか」を確かめます。

設定	最初の1本	慣れてから
大きさ（縦長）	576×1024	704×1280
長さ	5秒	6〜10秒
ステップ	8（速くする部品を使う）	8〜12
かかった時間（私の環境）	**約3.5分**	6秒で約8分・10秒で約14分

1. 動かしたい**画像を1枚**読み込みます（人物の写真風の絵など）
2. **どう動かすかを英語で1〜2文**書きます。例: `She stirs the pan with a wooden spatula and looks down at it.`
3. カメラを止めたいときは、後ろに次の文を足します: `A single continuous shot, no cuts. Camera: completely static, locked-off tripod shot, no zoom, no push-in, no pan.`
4. 声や音楽が要らないときは、さらに次の文を足します（あとで別の声を当てる場合）: `Audio: quiet room tone only, no dialogue, no voice, no music. No captions, no subtitles, no on-screen text.`
5. 生成を押し、できた動画を見て、人物・顔・服が崩れていないかを確かめます

🔴 **縦と横の大きさは、32で割り切れる数にします。**720×1280 を指定すると途中で止まりました。704×1280 なら通ります。

---

8. 日本語のセリフを話させる

MiniMax H3 は、日本語のセリフを口の動きに合わせて話せます。私の試しでは、できた動画の声を文字に起こすと、台本と一字一句同じ文になりました。

頼み方は、公式の書き方の形に合わせます。次の例の【 】の所を書き換えて使ってください。

```
For the target video, at 0.00 seconds into the target video, <Picture 1> (from [Shot 1]) is fully referenced.
```

```
integrated_multimodal_description:
[Shot 1] A static, locked-off eye-level shot. 【人物の見た目】 (S1, on screen, 【声の特徴：例
warm medium-pitched female voice, natural conversational pace, standard Tokyo Japanese accen
t】) 【どこを向いて、どんな表情か】. The woman (S1) says: <d>[Japanese] 【日本語のセリフ】</d>
【話し終わったあとの動き】. The camera does not move, no zoom, no cuts.

overall_soundscape:
【まわりの音：例 Quiet indoor room ambience.】

non_diegetic_music:
N/A
```

- セリフは `<d>[Japanese] …</d>` の中に日本語で書きます
- 話す人には (S1) のような印を付け、声の特徴を英語で添えます。同じ声の書き方にしておくと、別のカットでも同じ人の声に聞こえやすくなります
- **1本は15秒まで**です。長いセリフは、質問ごと・文ごとに分けて作ります
- 公式の解説では、声は8ステップだといちばん弱く、**12ステップ以上で安定する**とされています。声が崩れるときはステップを上げてください

---

9. 止まらずに流すための設定

- **生成の前に、ブラウザなどほかのアプリを閉じます。**生成中はメインメモリの空きが1GB 前後まで下がりました
- **使い終わった ComfyUI は止めます。**作り終わったあとも、メモリを確保したまま残っていました（止めると空きが約16GB 戻りました）
- **画像を作る作業や動画の書き出しと、同時に流しません**
- まとめて何本も作るときは、寝ている間に1本ずつ順に流すのが確実です（6秒の動画で1本約8分）

---

10. つまずきやすいところ

起きたこと	直し方
途中で止まる・パソコンが重くなる	ほかのアプリを閉じる。長い動画（10秒以上）ほどメモリを使うので、短く分けて作る

起きたこと	直し方
「shape … is invalid」で止まる	縦横の大きさを32で割り切れる数にする（720 → 704）
スマホやタブレットの画面の文字が、読めない英語のような文字になる	画面を見せるカットは動かさず、止め絵（画像のまま）で使う
頼んでいないのに、途中で顔に寄る・人を追いかける	座って手元を動かす場面はほぼ止まったまま。歩く場面は人を追いかけやすい。カットの切り替えに見えるならそのまま使う
テンプレートが一覧に無い	ComfyUI を 0.30.0 以上に更新する。起動のときの読み込みのエラーを見る
声が不自然	ステップを12以上に上げる。声の特徴の書き方をそろえる

---

11. かかる時間とお金の目安

項目	目安（私の環境）
モデルのダウンロード	約42GB（回線によって数十分〜）
縦長 576×1024・5秒	約3.5分
縦長 704×1280・6秒	約8分
縦長 704×1280・10秒	約14分
横長 1024×576・8秒（日本語のセリフつき）	約5.5分
お金	API 代は0円。かかるのは電気代だけ

広告動画1本には、場面ごとに7〜10カット使うことが多いです。夜のうちに流しておくと、朝にはそろっています。

株式会社スターストリーム・スタジオ この手順書は、ご自身で使う範囲でご利用ください。再配布・販売はご遠慮ください。