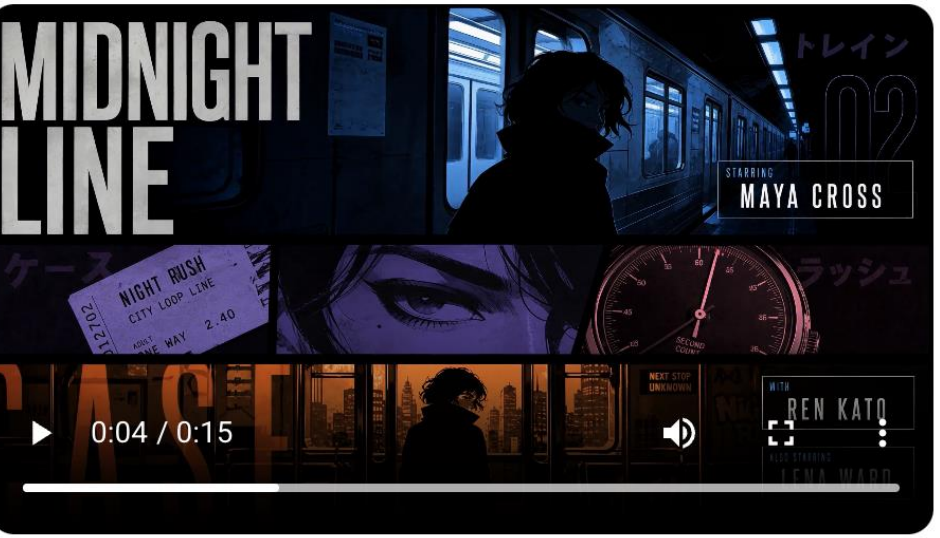


内田勉（ uchida @ mazariba.co.jp ）2026/9/6 自由研究発表会

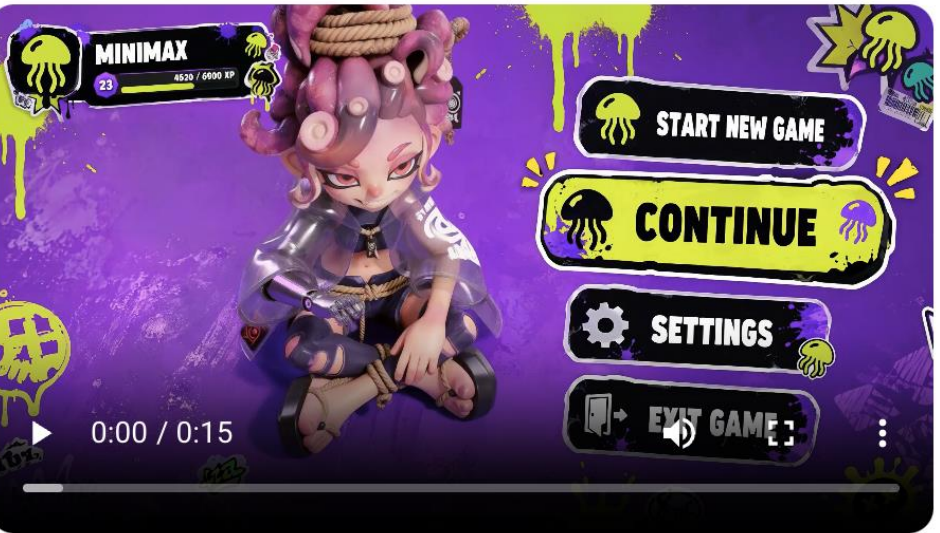
MINIMAX

H3 Use Cases

- Film Opening Titles



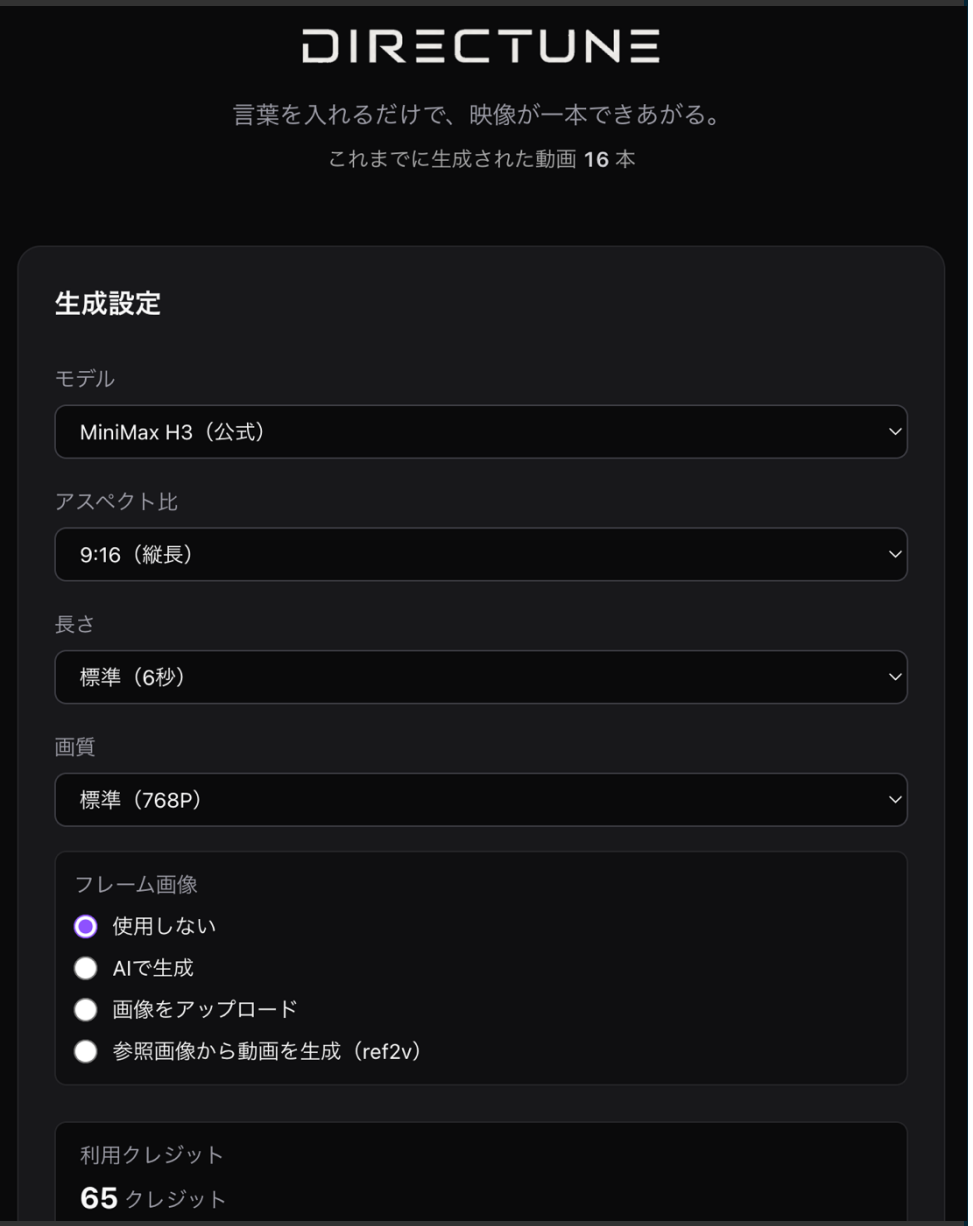
- Product Website



1か月前、動画生成AIに革命が起きた！

- 無料で使える MiniMax-H3登場7月31日
  - 最大15秒の長尺生成
  - コミュニティが色々改造
  - 日々、新機能、高速化が進む

- 俺の反応
  - うお～！ 無料で使える！
  - DgxSpark互換機(80万円)も買ったばかり！
  - やるしかねぇ！



動画生成サイトを作り始める

- 世の中には既に凄いクオリティの動画が溢れてる(世界規模で)
- 同じ土俵で戦うのはやめたい
- 自分なりの動画生成サイトを作り、使い勝手を改善しよう
- 作り始める



10日前 動画AIに再革命

- MiniMax-H3 MAX登場 8月27日
  - 動画の視聴時間より生成時間が短くなる
  - 5秒の動画を3秒未満、15秒の動画を約9～13秒で生成
  - 見終わる前に、次の動画が完成し、事実上の「無限長尺動画」が生成可能に

- 俺の反応！
  - すごい！　これは何かが起きる！



無限ギャルゲー登場

- 9月1日　X民に衝撃が走った！
  - https://x.com/LerSentAI/status/2094579488407302229



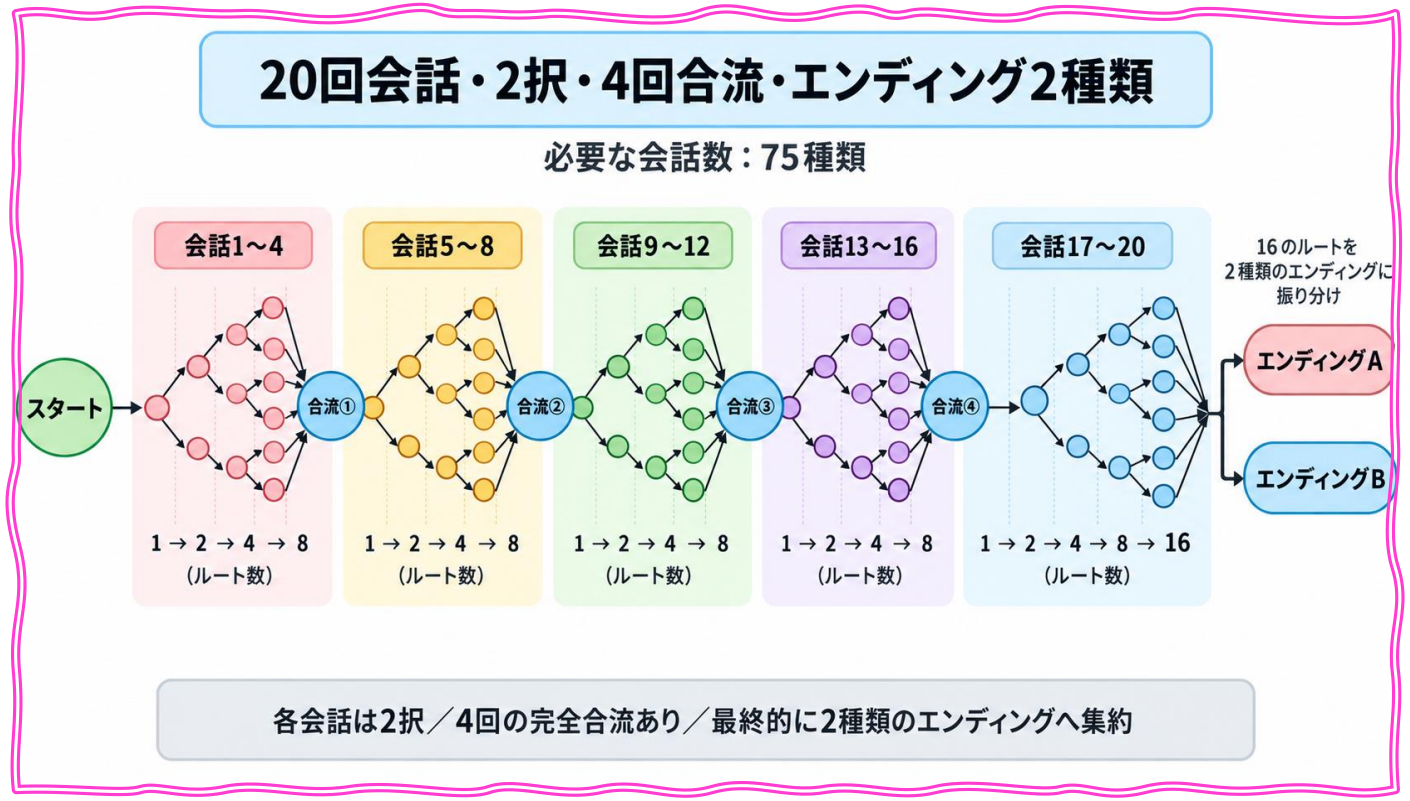


ギャルゲーを作り始める

- 当に動画生成すると大変だから、まずは静止画で作ろう
  - 調べると、会話数が爆発する事が判明
  - 20ターンの会話で34億8,678万通り
- ChatGPTからアドバイス
  - 途中で分岐を合流するといいよ！
  - 77シーンまで削減



分岐イメージ


20回会話・2択・4回合流・エンディング2種類
必要な会話数：75種類
会話1〜4
会話5〜8
会話9〜12
会話13〜16
会話17〜20
スタート
合流①
合流②
合流③
合流④
1 → 2 → 4 → 8
(ルート数)
1 → 2 → 4 → 8
(ルート数)
1 → 2 → 4 → 8
(ルート数)
1 → 2 → 4 → 8
(ルート数)
1 → 2 → 4 → 8 → 16
(ルート数)
16のルートを
2種類のエンディングに
振り分け
エンディングA
エンディングB
各会話は2択／4回の完全合流あり／最終的に2種類のエンディングへ集約



SHORT ROMANCE GAME

ふたりきりの部屋

付き合い始めたばかりの彼女と、初めて二人きりで過ごす午後。 言葉の奥にある気持ちを読みながら、会話を続けてください。

はじめる →

選択によっては、会話が途中で終わります。

BGM: MusMus「Pops up the mind wings」

3時間で静止画ギャルゲー完成

- プログラム、キャラデザ、シナリオ作成インフラ込で3時間で完成
- https://futarikiri-room.sidodtv.chatgpt.site/



62人が遊んでくれた！

全期間の概要

最終記録：2026/9/6 10:36

TOP訪問者	ゲーム開始
62	85
74セッション	40人が開始

ハッピーエンド	ゲームオーバー
28	27
開始比 32.9%	36回リトライ

DIRECTUNE

言葉を入れるだけで、映像が一本できあがる。

これまでに生成された動画 **16** 本

生成設定

モデル

MiniMax H3（公式）

アスペクト比

9:16（縦長）

長さ

標準（6秒）

画質

標準（768P）

フレーム画像

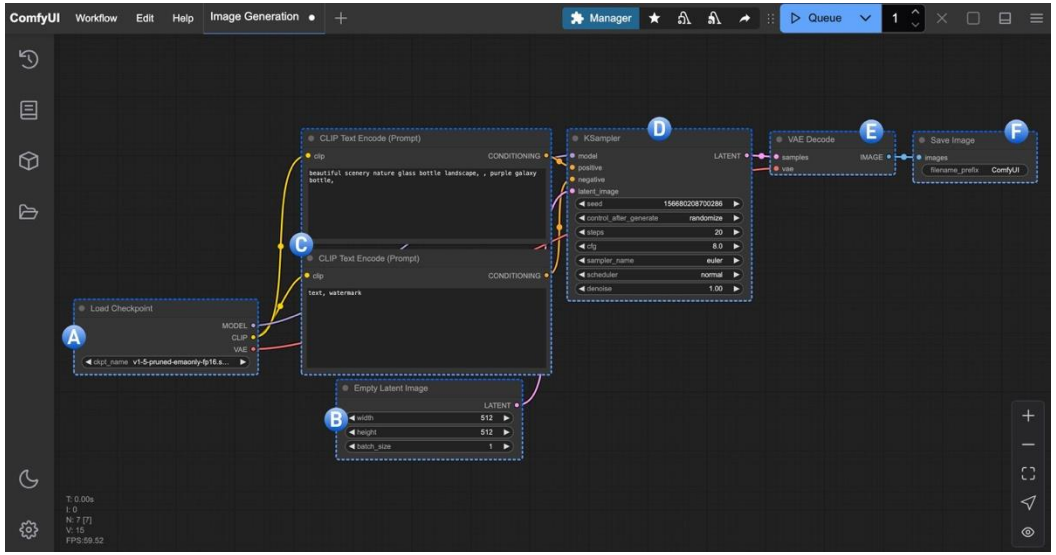
- 使用しない
- AIで生成
- 画像をアップロード
- 参照画像から動画を生成（ref2v）

利用クレジット

65 クレジット

いよいよ動画版を作るぞ！

- 会話数を減らす(エンディングは増やす)
- 意外と自作ツールが役立った
  - comfyUIが難しい





作って気づいたこと

- 自宅PC→有料APIへ(全28動画)
- ガチャが凄い
- 背景固定やクチパクが難しい
- 公式プロンプト作成Skillが凄く役立つ

- https://futarikiri-room-v2.sidodtv.chatgpt.site/



プロンプト全文

- 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] 2D-animated, cel-shaded anime style with clean line art. The vertical shot begins exactly as Picture 1: a young woman's head and shoulders fill the tall frame as she faces the lens in her own small room on a Sunday afternoon, her eyes open and resting warmly on the lens, a soft closed-lipped smile on her mouth, and a light pink colour already in her cheeks. She is Japanese and every word she speaks is spoken aloud in the Japanese language; she never speaks English or any other language, and the audio is Japanese from her first syllable to her last. Whenever she speaks, her jaw drops and lifts again on every single syllable and her lips change shape between wide, narrow and rounded from one syllable to the next, so her mouth opens and closes over and over through every line and is never held open in one fixed shape; whenever she is not speaking her mouth is closed. Her design stays exactly as in Picture 1 throughout: long straight dark brown hair with a fringe, the ends of the long strands in front of her shoulders fading from dark brown into blue over roughly their lower third in a soft gradient with no hard line, the blue only at the ends and never on the inner or under layer, never near the roots and never in the fringe, blue eyes, a white long-sleeved button-up shirt with a chest pocket, and a red necktie at her collar. These colours never change, never fade and never spread. This whole video is one fixed tripod shot: the camera never pushes in, never pulls back, never pans and never tilts. When she becomes larger in the frame it is only because she leans in toward the lens on her own feet; the camera itself does not move at any point. The room behind her is exactly the room in Picture 1, with nothing invented and nothing added: to the left the light natural-wood bed with pink bedding, the pink rabbit and beige teddy bear soft toys, the headboard shelf of small framed photographs and a little plant, and the framed pastel picture on the cream wall; to the right the light natural-wood desk with its white articulated lamp, pen cups of coloured pens, a pink cup and small white boxes, the open wooden shelving of books, storage boxes, plants in pink pots, framed photographs and a small black cat ornament, and the white shell chair. The daylight is soft, bright and evenly diffused with no harsh shadows and no hot spots, and the overall brightness, exposure and colour never change from the first frame to the last. Nothing behind her moves, shifts, appears, disappears, changes shape or changes colour, and the two soft toys do not move at all. For the first two tenths of a second the frame holds Picture 1 completely still, with no movement, no expression change and her mouth closed. Then her shoulders rise and fall once with a small breath and her smile warms. The young woman with a light, soft voice (S1), warmly and a little shyly, says on screen in Japanese, articulating every syllable with her jaw and lips and never holding her mouth open: <d>[Japanese] **私、君がこの部屋に来てくれて、本当に嬉しい。** </d> Her mouth closes on the last syllable. Her smile settles into something more serious and determined, her eyebrows draw in slightly, and the colour in her cheeks deepens as she holds the lens with her eyes. (S1) says in Japanese, more quietly and more firmly now, again articulating every syllable with her jaw and lips and never holding her mouth open: <d>[Japanese] **言葉だけじゃ足りないから、今日はちゃんと伝える。** </d> Her mouth closes. She lets a moment of silence sit, staying exactly where she is and exactly the same size in the frame, without leaning in at all and without coming any closer to the lens. (S1) says in Japanese, almost in a whisper, again articulating every syllable with her jaw and lips and never holding her mouth open: <d>[Japanese] **目、閉じて。** </d> Her mouth closes and stays closed for the rest of the video. She then holds completely still for a full second before she moves at all, and through this pause she stays exactly the same size in the frame: she does not lean in, she does not come any closer to the lens, her head does not move forward, and her eyes stay open and resting on the lens. The only changes during this held second are one slow blink, a small quiet rise and fall of her shoulders as she breathes, and her eyebrows drawing in a little as she gathers herself. Only once that full second of stillness has passed does she begin to move. Then, on her own and unhurried, she leans her face in toward the lens over about two seconds, filling more and more of the tall frame, while her eyelids come down little by little until her eyes are fully closed and her long dark eyelashes rest on her cheeks, the pink colour strong across them. She comes to rest close to the lens with her eyes closed, her lips together in a soft, calm line, her chin tipped very slightly up, and she stays there completely still and waiting, breathing quietly, without moving any closer and without pulling back. For the final two tenths of a second the frame holds completely still, with no movement, no expression change, no eyelid movement and no mouth motion.

- overall_soundscape: No room tone or outdoor ambience; only close-microphone human sound. The opening and closing two tenths of a second carry no sound at all. Her voice is close and intimate at the microphone throughout, dropping to a near-whisper on her third line. A small breath opens the shot and light breathing sits between the lines. After her last line there is no speech at all: first a full second of near silence in which only her quiet breathing is heard, with no rustle of hair or clothing at all because she is not yet moving; then a soft rustle of hair and collar begins as she starts to lean in, and her quiet, steady breathing grows a little closer to the microphone as she comes nearer and then holds. There are no footsteps at any point.



作って気づいたこと

- 1カット15秒だと、案外足りない
- 2カットを連結するのが大変
  - 連結の際にUP Scaleが必要
  - UP Scaleが高くて遅い

- 彼女の目線が冷たい。。。



良かったら遊んで感想を下さい！
登録不要・無料です！

動画版



半分ぐらいまで
出来てます

静止画版



内田勉のX

内田勉のブログ