

Kataribe: Audiobook.AI User's Manual Beta V1.0

Kataribe: Audiobook.AI (語り部, "storyteller") is a narrated video maker that turns a novel manuscript into **narrated videos, one per chapter** (portrait or landscape). Writing the script → generating audio and images → checking → building the video are all done on this single screen. Generation uses the APIs of external services, and you pay each service directly. At minimum you need two, ElevenLabs and Gemini (or OpenAI); at most four, adding Anthropic and HeyGen.

Contents

1. [Installation and setup](#)
2. [Production flow \(overview\)](#)
3. [New project and chapter list](#)
4. [Script \(human gate 1\)](#)
5. [Generating audio and images](#)
6. [Check \(human gate 2\) and how to fix](#)
7. [Greeting, closing and BGM](#)
8. [Build and finished files](#)
9. [Settings reference](#)
10. [Cost estimates](#)
11. [Troubleshooting \(FAQ\)](#)
12. [Glossary](#)
13. [Terms \(EULA\)](#)

1. Installation and setup

1. There are two ways to get it. Neither needs administrator rights, and there is no need to install Python or ffmpeg (they are bundled).
 - **Microsoft Store (recommended)**: click the "Get it from Microsoft" badge on the website, or search for "Kataribe" in the Store app and click Get (<https://apps.microsoft.com/detail/9nph4438c3cl>). Updates are delivered automatically by the Store. Uninstall by right-clicking it in the Start menu or from Settings → Apps. The Store edition keeps the API keys (.env), voice dictionaries and log in %LOCALAPPDATA%\Kataribe ("About this app" shows the location)
 - **Installer (Setup.exe)**: run the exe downloaded from the website (such as Kataribe_Beta_V1.04_Setup.exe) and follow the wizard. It installs into %LOCALAPPDATA%\Kataribe and creates a "Kataribe" shortcut on the desktop and in the Start menu. Uninstall from Windows Settings → Apps (a backup of the API keys (.env) and voice dictionaries is kept in Documents\Kataribe and restored automatically by the next install). Setup.exe, the bundled ffmpeg and the libraries are all signed with RAY Inc.'s certificate, so it runs even on PCs where Windows 11 Smart App Control (SAC) is "On". While downloading, your browser may say the file "isn't commonly downloaded" and leave it named "Unconfirmed crdownload" (a new release has few downloads yet; the file itself is fine). Open the downloads list (Ctrl + J in Edge) and choose "... " → "Keep" → "Keep anyway". If "Windows protected your PC" appears the first time you run it, click "More info", check that the publisher is "株式会社RAY" (RAY Inc.), then click "Run anyway". If the publisher is different, do not run it and contact us
2. Launch Kataribe → a dedicated window opens (no browser is used). Close it with "Quit" at the top right of the screen
3. On first launch, Home shows "Start first-time setup" (register keys → connection test → first project). You can change these later under "Setup"
4. **Pro edition**: enter the key you purchased (KTRB-XXXX-XXXX-XXXX-XXXX) under "Pro license key" in Setup to remove the watermark from finished videos, use multi-actor (actors), and use the app commercially. The free edition has every production feature (with watermark, no commercial use; limits are 3 minutes per chapter, 3

chapters per project and 5 finished chapters per month; the closing is a fixed 3-second "Powered by Kataribe: Audiobook.AI" card). "About this app" shows your current edition. Commercial use also requires a paid plan on the ElevenLabs side

5. **Tool language:** switch between 日本語 (Japanese) and English under "Tool language" in Setup (documents such as the manual and the terms switch with the language too)
6. "Try the sample first" → "Create sample (Japanese / English)" on Home creates a project from the bundled short story (2 chapters, about 1 minute each). We recommend making your first video with it to experience the flow (the same button appears at the end of first-time setup)

Service	Used for	Required?	Where to get it
ElevenLabs	Narration audio, BGM	Required	elevenlabs.io → Profile → API Keys
Gemini	Story images, script drafts	Either this or OpenAI is required	aistudio.google.com/apikey (image generation requires billing ; it does not work on the free tier)
OpenAI (ChatGPT)	Story images, script drafts	Either this or Gemini is required	platform.openai.com → API keys (add credit under Billing)
Anthropic	Let Claude draft scripts and auto-fix NG cuts	Optional (otherwise the image AI does it)	console.anthropic.com → API Keys
HeyGen	Lip-synced narrator avatar video	Optional (only for the avatar structure)	API dashboard at app.heygen.com

Keys are stored only in `%LOCALAPPDATA%\Kataribe¥.env` on this PC and are sent nowhere except to authenticate with each service. The projects folder can also be changed in Setup (default: Documents\Kataribe\projects).

2. Production flow (overview)

New project → chapter list → create chapter folders → **Script (approval)** → generate audio and images → **Check (OK/NG)** → greeting, closing and BGM → Build → final mp4

Only the two bold steps need human judgment. Everything else proceeds with a single button, and "Next step" on each page tells you what to do now. The top of the chapter page has a progress bar (Script → Generate → Check → Build → Done) with the current stage shown in gold. The project's chapter list shows the same stage badges and the number of items remaining.

3. New project and chapter list

- "New project" on Home: choose the manuscript (.docx / .txt / .pdf), the title and its reading, the narrator's name, the voice (pick from the list; ► Preview lets you listen), the aspect ratio (9:16 portrait / 16:9 landscape / 1:1), the target length per chapter, and the chapter structure (greeting, closing, logo, header)
- Selecting a manuscript immediately shows a chapter preview (number of chapters, headings, character counts) and the detected language. With PDFs, vertical writing or multi-column layouts can scramble the text order, so verify with "Review / fix chapters" after creating the project (scanned image-only PDFs yield no text)
- Chapters are detected automatically from 「第N章」 (Chapter N) headings in the manuscript. For other styles ("1-1.", "Chapter 1", etc.), use "Review / fix chapters" on the project page to pick from the candidate buttons or specify a regular expression
- For manuscripts whose chapters are only a few lines each (story collections, poems), use "Treat the whole manuscript as one chapter" (checkbox on New project, or "Review / fix chapters"). The text is not split: heading lines stay in the text and are narrated as their own cuts (only the first heading gets a picture; later headings carry over the previous one), and the chapter takes the work title. Switching it later archives the existing chapter folders to the trash so you can recreate them
- Selecting a manuscript fills in the project name and title from the file name; for .txt it also detects the language (Japanese / English) and whether chapter headings exist, and guides you. A logo image (png / jpg) can optionally be set here too
- "Create folder" creates a working folder for each chapter (inside the projects folder). For chapters that already have a folder, this column shows the last-updated time instead (newer of the chapter state and log)
- The image style can be picked from a preset (Cinematic / Watercolor / Sumi-e ink / Anime background / Oil painting / Picture book / Photorealistic / Film noir / Sci-fi) and edited by hand

Presentation (PPTX) mode (V1.04)

- Choose a **.pptx** as the manuscript in New project and the project runs in presentation mode: each slide becomes the picture of its cut, and the text in each slide's **speaker notes** is narrated as written. No AI writes the script

- Long notes are split at sentence boundaries into cuts of about 100 characters (they keep the same slide picture). Slides without notes are skipped
- Slides are exported to PNG with the PowerPoint installed on this PC (into スライド/ under the project). On a PC without PowerPoint, export the slides from PowerPoint with File → Export → PNG, put them in スライド/ as slide001.png ... and press "Re-import slide pictures" on the chapter page
- Defaults at creation: the aspect ratio follows the slides (e.g. 16:9), opening, closing and the header are off, and narration captions are not shown (enable them with "Narration captions (presentation)" in Settings)
- The flow is the same as for a novel: create the folder → review and approve the narration on the script page → generate audio → check → build. Pictures are not AI-generated, so replace an NG picture by uploading one
- Stage directions in parentheses in the notes (e.g. “(lower the voice, sound serious)”) are not read: they are removed and an emotion tag such as [serious] is placed before the next sentence. Ordinary parenthetical notes are read as usual
- A PPTX without speaker notes works too. For slides without notes the AI writes the speech from the source document if there is one, otherwise from the slide's title and body (automatically when the folder is created; upload a source document later and press “Write speech and rebuild script” to redo it). Without a script-AI key, slides without notes are skipped
- **Write the speech with AI:** “Write speech and rebuild script” on the chapter page writes the speech for slides without notes (or all slides) in four stages. ① Read the source document (optional; upload .docx / .txt / .pdf under “Source document”) and organise its key points → ② grasp each slide's content (image-only slides are transcribed by the AI from the picture) → ③ write, per slide, what to say about what is on that slide in the order it appears (the slide comes first, the document second; if they disagree the slide wins) → ④ check each speech against its slide and rewrite only the ones that drifted. “Tone of voice” (e.g. bright and lively / calm and polite; also in Settings) shapes the wording and emotion tags. The result is kept in 原稿/AIスピーチ.json (with the digest and transcriptions) and can be edited on the script page
- After editing the PPTX, replace the .pptx in 原稿/ and press "Re-import slide pictures" (only the slides changed) or "Rebuild script from notes" (the notes changed too; audio is remade) on the chapter page

4. Script (human gate 1)

- Chapter page → "Edit / approve script". Creating a chapter folder starts drafting the script automatically (multiple chapters run in order, 1–2 minutes each). To redo it, press "(AI name): re-draft the script". Cuts, narration text and scenes (image prompts) are created automatically from the source text. Which AI does this is chosen under "Generative AI" in the project settings (images: Gemini or ChatGPT; script: the same AI as images, or Claude)
- Rules: narration text per cut is **60–100 characters** (longer cuts fall apart towards the end), the total is within $\pm 15\%$ of the target length, numbers are written in kanji numerals, and emotion tags such as [softly] are used sparingly. Do not start a cut with a Sino-Japanese compound (two or more kanji) or a proper noun (pronunciation breaks easily; put a lead-in word such as 「その」 (sono, "that") or 「やがて」 (yagate, "before long") in front). The script page warns about violations of these rules
- The caption text is created automatically from the narration text ("Rebuild captions from narration"). If narration and caption diverge, caption timing goes wrong too, so the trick is not to change it much by hand. Captions are built against the chapter's source text, so words the narration spells out in kana for pronunciation (names, hard words, Latin letters such as AI) return to the source spelling. A cut whose caption spells a character's name inconsistently gets a warning on the script page (names registered as actors)
- To leave a cut without captions, set its "Captions" to "No captions" and save (there are buttons to switch all cuts at once). The caption text is kept but not burned into the video. Rebuild chapters that were already built
- The labels 挨拶 (greeting) and 締め (closing) are reserved for the avatar bookend cuts. A body cut with one of these labels is left out of the video, so use another label (the script page warns you)
- Images are generated only for cuts that have a "scene". If you save with the scene empty, that cut's image is archived and the video carries over the previous cut's picture. To give every cut an image, use "Fill empty scenes with Claude"
- Speaker: once actors are registered in Settings, each cut lets you choose a "Speaker" (the narrator or an actor's name). Drafting also assigns characters' lines to actor cuts (an actor cut's narration text is the line itself, without 「」 quotation marks or

"he said"). An actor cut's caption is shown automatically as "Name 「line」" so viewers can tell who is speaking (the audio reads only the line).

- World bible: "World bible (shared image setting, English)" in Settings is prepended to every image prompt to pin the setting, era, materials and light. If empty it is generated from the manuscript at the first draft; "Generate from the manuscript with AI" regenerates it. When off-world pictures appear in an existing chapter, use "Rewrite scenes to match the world bible" on the Script page, then regenerate the images
- Locations and reference images: so that pictures of the same place do not turn into different rooms from shot to shot, each cut in the script has a "Location" (e.g. Osaka event hall), and a set design per location (layout, placement of furniture and fixtures, materials, light) is added to every image prompt. Pictures already made for the same location are also sent to the AI as reference images, so a new camera angle keeps the same set. Each cut also gets a time of day (dawn to late night), and the view through windows, the sky and the light follow it (a continuous scene keeps the same time). Locations are assigned by the AI after drafting, and for older chapters when their images are made. Scenes in a new draft are written to match the set designs decided in earlier chapters, and a different character's workplace or room is designed as a separate location (one organisation's office shared by several characters stays one location). Edit them with "Location" and "Time of day" on the Script page and "Location designs" in Settings. To turn this off, set "Keep pictures of the same location consistent" to No
- Acting for actor lines: drafting and speaker assignment prefix each line with delivery cues read from the surrounding narration (emotion tags such as [whispers] [shouts] [sighs] and pauses with "..."). Existing chapters get them via "Split dialogue from narration and reassign speakers" If you change the speaker and save, that audio is archived because it belongs to a different voice and must be regenerated. Lines of characters not registered as actors also become separate cuts with the speaker name attached (with no voice set, the narrator's voice is used; add the character as an actor in Settings to give them a voice). When dialogue and narration are mixed in the same cut a warning appears; split them with "Split dialogue from narration and reassign speakers" (cut numbers change, so audio and images are archived and the script must be approved again)

- After reviewing the content, press **"Save and approve"**. The generate buttons cannot be pressed until the script is approved
- A cut inserted with "+ Add cut below" gets a suffix number such as 26b, 26c (later cuts are not renumbered: audio and image file names are derived from the number, so renumbering would force regeneration). An empty number is filled in automatically on save, and duplicate numbers are rejected

5. Generating audio and images

- "Generate audio" and "Generate images" on the chapter page. The buttons show the number of missing items and the estimated cost; pressing one opens a confirmation dialog
- Generation runs **4 at a time with no pause** (when a rate limit is hit, it waits and retries automatically). Progress appears in the Working bar at the top as a count (audio 12 / 48) and time remaining; the page refreshes automatically when done
- Existing items are not regenerated. To redo one, use the Check screen for that cut. To regenerate every cut's audio, use "Regenerate all audio" on the chapter page (current audio is archived to `_old/` and the checks are reset to unchecked)
- To stop midway, press "Stop" in the Working bar. Generation stops after the current 3 items finish and what was made remains (press "Generate" again to continue). A build stops at the current step and partial results carry over to the next run
- When a job that took 20 seconds or more finishes, a Windows notification sound plays and the screen refreshes automatically (so you notice even while working in another window)

6. Check (human gate 2) and how to fix

- Chapter page → "Check". For each cut, listen to the audio, look at the image, and mark **OK** / **NG**. For NG, write a memo (e.g. 「種がたねと読まれた」, "種 was read as tane")
- Filter the view at the top (All / Unchecked only / NG only / Flagged only). "Next unchecked" and the keyboard shortcuts (J/K next/previous, Space play, O OK, N NG) let you move fast. "Mark all cuts OK" and "Reset all cuts to unchecked" are also at the top
- **Flagged**: cuts whose audio length differs greatly from the estimate based on character count (possible skipped or repeated words) are flagged automatically. If it sounds fine, mark it OK
- **Reporting inappropriate AI content**: if a picture, audio or text made by the AI contains inappropriate material (discrimination, violence, sexual content, rights infringement and so on), report it to the developer from the top of the Check page or the "Report" link of each cut. Your mail app opens a report email filled in with the project, chapter and cut (the app never sends anything automatically). It can also be opened from About
- **Audio NG**: marking NG opens three ways to fix it. ① **Auto-fix** (based on the memo, tries the pronunciation dictionary's rules first and then AI, and **always changes the narration text** to make one new take, because regenerating the same text does not fix it), ② **Rewrite the narration yourself** (regenerates with your rewritten text), ③ **Compare reading candidates (advanced)** (write several spellings of a word, generate them all at once and compare by ear; useful for numbers and proper nouns). All three rebuild the caption from the new text (words opened into kana only for pronunciation keep their original spelling in the caption)
- New takes line up on the card. You can see "what was changed and how" (spelled in kana / rephrased / lead-in added) and the narration text at that time. Compare them and press "Use this take" → it is marked OK automatically and the script and caption are aligned to that text
- **Image NG**: write a memo (what is wrong, e.g. "a person is visible", "too dark") and press "Auto-fix"; the AI rewrites the scene to reflect the memo and then regenerates (it never regenerates the same scene; without a memo it changes composition, viewpoint and lighting). You can also write the scene yourself and regenerate. "Fix

NG images from memos and regenerate" does this in bulk. The new scene is recorded on the take and applied to the script when you use it

- **"Redo all NG cuts"** at the top of the Check page regenerates NG audio and NG images in one go (audio via dictionary fixes or rephrasing, images by rewriting the scene from the note). You still listen/look at the new takes and use them as before
- **Use my own image:** "Use my own image" in each cut's image section lets you pick a file; it is scaled and cropped to the video's aspect ratio and appears as a take (press "Use this image" to use it). If added to a cut without a scene, it is adopted as is
- If you change the narrator's or an actor's voice in Settings, audio for cuts using that voice is archived and returns to "not generated" (make it with the new voice via "Generate audio" on the chapter page).
- Actor cuts carry a name badge, and the audio section also shows the speaker (narrator / actor name). Fixing works the same as for narrator cuts (regeneration uses that actor's voice)
- Reading fixes confirmed by "Use" are recorded automatically in the work's pronunciation dictionary and applied from the start in the following chapters
- On the pronunciation dictionary page (from the project page), pick one of three types. **Fix the reading** (write the target as “word (correct reading)”, e.g. 値踏み (ねぶみ) ; script drafts and auto-fix open it into kana, captions keep the kanji), **Watch (note only)** (the reading is not replaced; the word is only marked as a watch word) and **No problem** (removes the watch mark). When in doubt, “Fix the reading” is all you need. Chapter, Symptom and Fix may be left blank (Symptom is matched against NG notes). Entering the main characters' names and readings before you start a work helps the most

7. Greeting, closing and BGM

- **Greeting and closing:** the "Opening and closing" card on the chapter page. Only what the configured mode (default: title card) needs is shown. For "Avatar": "Generate opening greeting video" and "Generate closing greeting video" (a lip-synced video that consumes credits if HeyGen is configured; otherwise a still video of the standing image). For "Title card": a silent still card (generated automatically at build time). For "None": just a note saying so. The background is set under "Title card background" on the chapter page: from a scene description (in English; pressing "Let AI read this chapter and write the scene" grasps the chapter's setting, motifs, emotional arc and intent, writes a symbolic scene, and shows its reading below the field), generate as many candidates as you like, then pick one from the thumbnails (click to enlarge) with "Use this background". You can also import your own images as candidates
- Lip-sync material is saved, and from the second time on only the captions are rebuilt (free). If you regenerate the greeting audio and adopt it, the video is rebuilt automatically; to force a rebuild, use "Rebuild from the avatar"
- **BGM:** the BGM section on the chapter page. The music prompt (in English) is written automatically by AI from the chapter's content (redo with "Let AI write the music prompt from the chapter"). Generate from that prompt (consumes ElevenLabs credits; 75 seconds $\approx$ 5,000) or import your own mp3 etc. "Use this BGM" mixes it softly under the story at build time (volume based on the setting, 0.15 by default; the track loops with fades). "Preview with narration" on each candidate plays the first 25 seconds of the chapter's narration with the BGM at the configured volume (to check before adopting; the same mix as the build)

8. Build and finished files

- "Build" on the chapter page. Audio conditioning → caption layout → video rendering → waveform verification → header burn-in → BGM → joining greeting + story + closing, all automatic (no charges; a few minutes to ten-odd minutes. With the title card structure, the opening and closing still cards are also generated here automatically)
- Output: <title>_第N章_配信用.mp4 (the final video for distribution, with greeting and closing) and ..._本編.mp4 (the main part, narration only) in the chapter folder. A completion card appears at the top of the chapter page with Play / Open folder / Main part only / Frame check image. The video player page has a seek bar to move the playback position
- The bottom left of the finished video carries a small "Powered by Kataribe" (free edition; the Pro edition can turn it off in Settings). The mp4 metadata records the edition (Free / Pro license ID)
- Fixing captions or the logo and rebuilding is free, as many times as you like. Generated files are never deleted; they are archived to _old/
- To start a chapter over, use "Clear everything and restart" at the bottom of the chapter page (generated files are archived to _old/)
- **Trash and archives:** "Trash" in the footer lists deleted projects (Restore / Delete permanently) and each chapter's archive folders (_old) (Open folder / Restore to chapter, which puts back the state before a full reset / Delete permanently). To move a project to the Trash, use "Delete" in the Home list or "Delete (to trash)" at the bottom of the project page. The project name (folder name) can be changed with "Rename project" at the bottom of the Settings page (separate from the displayed title). If you change the work title in Settings, rebuild each chapter: the video header and file names then use the new title (old-title videos are moved to _old/題名変更 at build time)
- The finished video is encoded with settings tuned for still pictures (V1.04), so slide and scene-picture videos stay small at the same quality (about 10 MB for a 2 min 34 s presentation). Chapters built earlier shrink as their cuts are re-rendered on the next build

9. Settings reference (project page → "Settings")

The Settings page shows every item. "Hide advanced items" at the top right hides AI model names, reading speed, notation rules, templates and so on (remembered next time)

Item	Description
Title, reading, brand name, narrator	Inserted into the greeting template ({narrator} {novel_reading} {brand} {chapter_kanji} etc.)
Voice	From the ElevenLabs list. Each voice's pronunciation quirks accumulate in the dictionary under <code>voices/</code> and are reused across works
Voice model	The ElevenLabs text-to-speech model (eleven_v3, eleven_v4, etc.; the list comes from the API). Default: eleven_v4. v3 or later (emotion tags) is recommended. A change applies to audio generated from then on
Actors (multi-actor)	Pro edition. "Detect characters from the manuscript" fills in just the names of named characters with dialogue (also automatic on project creation). Then simply pick a voice for the characters you use. Up to 12 characters, each with the name used in the story and an ElevenLabs voice. Only rows with a name count. A blank voice means the same voice as the narrator. The actor voices' pronunciation dictionaries are applied to the narration too
HeyGen avatar, type, engine	From the first 40 of your photo avatars. For a LOOK ID not in the list, use "Enter manually..."
Standing image	Still avatar when HeyGen is not set / title card background
Aspect ratio, target length, reading speed, BGM volume	Video canvas (9:16=1080×1920 / 16:9=1920×1080 / 1:1), the basis for script length, BGM ratio
Chapter structure	Whether to include a greeting / closing, their form (avatar / title card), logo, top-right header (logo, title, chapter name; when the logo is hidden, title and chapter name align to the right edge), handling of parts (第N部, Part N)

Caption size, font, weight	Small / Normal / Large / Extra large. Font from the 4 bundled ones (Mincho: Noto Serif JP, Shippori Mincho; Gothic: BIZ UDPGothic; rounded Gothic: Zen Maru Gothic), bold or not. Header, greeting and title card text use the same font. Changing these re-renders the video on the next build
----------------------------	---

Caption notation rules	Numerals (kanji / Arabic half-width / full-width), Latin letters. Narration always uses kanji numerals for pronunciation reasons
------------------------	---

"Powered by Kataribe" at bottom-left	Pro edition can choose Hide / Show (default: hide). Always on in the free edition
--------------------------------------	--

Image style, greeting template, closing template	The art style shared by all cuts, and the fixed wording of the greeting and closing
--	---

Choose a file under "Logo image" on the Settings page (or place `logo.png` / `logo.jpg` directly in the project folder) and the logo appears at the top right of the greeting, closing and story.

10. Cost estimates (measured)

Item	Estimate
Narration audio	About \$0.0002 per character (a 6-minute chapter $\approx$ 1,800 characters $\approx$ \$0.4). Also consumes ElevenLabs' monthly character quota
Images	About \$0.04 per image (20 images $\approx$ \$0.8)
BGM	About 5,000 ElevenLabs credits per 75 seconds (shared with the character quota)
HeyGen lip-sync	Credits for the avatar's duration. Greeting + closing (about 30 seconds) $\approx$ \$1
Script drafting and auto-fix (same AI as images, or Claude)	A few yen to a few tens of yen per chapter

A complete 6-minute chapter costs **roughly \$1.5–3** (depending on how many times you regenerate). The generate buttons always show an on-the-spot estimate. The project page shows "Generated so far $\approx$ \$" (estimated from the number of takes) and this month's ElevenLabs character usage. The card on the chapter page shows that chapter's share

11. Troubleshooting (FAQ)

The audio reading does not get fixed

Regenerating the same text does not fix it. Change the text in one of these ways: spell it in kana (種→しゅ), rephrase it (懷疑→疑い), or put 「その」 in front. "Auto-fix" does this automatically and shows what it changed on the take.

A cut was marked "Flagged"

The audio length differs greatly from the estimate based on character count (possible skipped or repeated words). If it sounds fine, mark it OK; if not, mark it NG and press "Auto-fix".

The narration falls apart or drops lines towards the end

The cut is too long (it tends to happen above 120 characters). Split the cut at a sentence break in the script. Automatic drafts keep cuts to 60–100 characters.

Captions and narration are out of sync

Rewriting the caption text by hand breaks the display timing, which is decided by character-count ratio. Realign them with "Rebuild captions from narration" on the script page.

Text or people appear in images

Remove descriptions of text, signs and people from the scene (English) and regenerate. Also check that the shared image style includes "No text, no people".

Gemini image generation fails with 429

The free tier has no image quota. Enable billing via "Set up billing" at aistudio.google.com/apikey. You can verify with "Generate one test image" in Setup.

A button cannot be pressed

Either the script is not yet approved (generate buttons) or a job is running. While a job is running, a "🕒 Working" panel is shown, and the buttons re-enable automatically when

it finishes.

Cannot go back from a video in the dedicated window / cannot select text

Videos and images open in the app's own player page; go back with "← Back" at the top left. Text can be selected and copied. If the window is blank, try `Kataribe_Browser.cmd` (fallback).

Updating the app

The Microsoft Store edition is updated by the Store automatically (to update right away, open the Store app → Library → Get updates). For the installer edition, just download the new setup exe from the website and run it. In both cases API keys (.env) and projects are kept as they are. The setup closes a running Kataribe for you.

Where are the logs?

Each chapter's job log is at the bottom of the chapter page. The app's own log is

```
%LOCALAPPDATA%\Kataribe\kataribe.log .
```

12. Glossary

Term	Meaning
Project	The container for one work. Holds the manuscript, settings, pronunciation dictionary and chapter folders
Chapter folder	The working place for one chapter. Holds the script, audio, images, takes and finished videos (compatible with the manual-era layout)
Script	The blueprint of the narration: the chapter split into cuts. Subject to gate 1 (approval)
Cut	The smallest unit of the script. Has narration text, caption text, scene and speaker; one audio file and one image correspond to it
Narration text (TTS)	The text actually read aloud. 60–100 characters, kanji numerals, emotion tags used sparingly
Caption text	The subtitles shown on screen. Made automatically from the narration text; blank lines separate blocks (screen changes)
Scene	The image prompt for that cut (English). If empty, the previous cut's picture carries over
Location	Where the cut's picture is set (e.g. Osaka event hall). Cuts at the same location are drawn with the same set design and reference images
Narrator / actor / speaker	The narrator is the voice that reads the narrative. Actors are per-character voices (Pro edition). The speaker is which of them reads a cut
Take	One generated audio file or image. Takes line up on the card; compare them and "Use" one
Use	Making that take the official audio or image. Using a take marks it OK and applies its narration text or scene to the script

Check (OK / NG)	Gate 2. Audio and images are judged separately, by listening and looking. Generating alone does not make it OK
Gate	The 2 places where a human decides: script approval and Check. These are never automated
Flagged	The mark put automatically on cuts whose audio length differs greatly from the estimate based on character count (possible skipped or repeated words)
Auto-fix	Reads the NG memo and always changes something before regenerating: a dictionary rule or an AI rephrase (for images, a rewritten scene)
Reading experiment	Write several spellings of a word, generate them all at once and compare by ear (③). For the readings of numbers and proper nouns
Pronunciation dictionary	The rules for readings. The work's dictionary (proper nouns etc.) and the voice's dictionary (the voice's quirks, carried across works)
Risky word	A word whose pronunciation breaks easily with that voice. Do not place it right after a tag
Opening / closing	Before and after the story. Title card (still, silent), avatar (the narrator speaks), or none. In the free edition the closing is fixed to the Powered by card
BGM candidate	A generated or imported track. Verify with "Preview with narration"; once used, it is mixed under the story at build time
Build	Audio conditioning → captions → video → verification → header → BGM → joining opening and closing. No charges, any number of times
Main part / final	The main part (本編) is the narration video without a header. The final (配信用) is the finished mp4 with opening, closing, header and BGM

_old (archive)	Where files displaced by regeneration or settings changes go. Nothing is deleted, so they can be restored
Trash	Where deleted projects go (projects/_trash). Moving the folder back restores the project
Free / Pro edition	The free edition has the watermark, the closing card, scale limits and no commercial use. The Pro edition is unlocked with a license key and allows commercial use
License key	16 characters, KTRB-XXXX-XXXX-XXXX-XXXX. Registered and removed in Setup

The "?" next to each screen's heading jumps to the relevant section of this manual.