Scribewave Airgap Institutional Evaluation Pack

Everything a security review, an ethics board, or a procurement office needs, in one document

Scribewave · hello@scribewave.com

airgap.scribewave.com/network

1. Summary

Scribewave Airgap is a desktop application that transcribes recordings, labels speakers, detects and redacts personal data, and answers questions about a document library. All of it on the computer it is installed on.

The distinguishing property for an institutional buyer is negative rather than positive: **there is no server-side processing.** No recording, transcript, document, filename or usage event is transmitted to Scribewave or to any third party. There is therefore no processing agreement to negotiate over content, no sub-processor chain to assess for content, and no international transfer to justify.

Product	Scribewave Airgap (bundle identifier `com.scribewave.offline`)
Type	Native desktop application — macOS and Windows
Where processing happens	Entirely on the end user's device
Content sent to the vendor	None
Analytics / telemetry	None. No analytics SDK is embedded.
Network dependency	Licence activation once; model download once; optional update check
Individual licence	€79 one-time, up to 3 devices, 12 months of updates
Institutional licence	Annual site licence, perpetual-fallback
Trial	7 days, full functionality, no payment details required

2. What Airgap is, and what it is not

It is a complete local workflow for recorded speech: batch ingestion of most media formats, an unattended processing queue, speaker diarization, a word-level transcript editor with waveform and click-to-seek, custom vocabulary applied before transcription, PII detection with redacted export, local meeting recording for Zoom, Teams and Meet, dictation, semantic and full-text search across the library, on-device AI summaries and chat, and export to TXT, DOCX, PDF, SRT, VTT, CSV, plus QDA-specific formats for NVivo and MaxQDA.

It is not the most accurate transcription product Scribewave makes. That is Scribewave online, the cloud sibling, which runs larger models on GPUs and reaches 2.3% word error rate. Airgap runs Whisper large-v3-turbo, at 4.4% in English and 5.9% in Dutch on FLEURS. §7.1 sets out what those numbers do and do not cover. We state this in our own sales material because an evaluation that discovers it later is an evaluation we deserve to lose. Airgap is the correct choice when the recording cannot be uploaded, or when a fixed one-time cost is preferable to a per-hour bill. It is not the choice when maximum accuracy is the primary requirement.

It is not a server product. There is no on-premise deployment, no administration console, and no central management interface. Each installation is independent.

---

3. Data protection posture

3.1 The core position

Under GDPR terminology, Scribewave does not act as a processor of the content your users put into Airgap, because Scribewave never receives that content. The recordings, transcripts, and any personal data within them are processed solely on the institution's own device, by software the institution has installed, under the institution's own controls.

The practical consequence that most shortens an evaluation: **for the content itself there is generally no data processing agreement to sign, because there is no processing by us to govern.** Institutions should confirm this reading with their own DPO, since we are describing the architecture and not offering a legal opinion. But the architecture is unusually simple to assess, and §4 exists so that assessment can be done empirically rather than on trust.

3.2 What Scribewave does receive

Exactly one category of data, and only from purchasers rather than from users of the software:

- **Licence administration.** When a licence key is activated, the application sends the key, a hashed device fingerprint, and the device name, so that the three-device limit can be enforced. Purchase and billing data is held as for any commercial transaction.

No content, no metadata about content, and no behavioural data is included in this or any other transmission.

3.3 Jurisdiction and international transfers

Scribewave is operated by **Quantum Noodle BV**, a company established in Belgium and therefore within the EU/EEA. The licence-administration endpoint described in §3.2 runs in AWS `eu-west-3` (Paris).

The consequence for a European institution: there is no third-country transfer to assess anywhere in this product. Not for content, since none is transmitted at all, and not for the licence-administration data either, which stays with an EU-established supplier on EU infrastructure. No standard contractual clauses, no transfer impact assessment, and no US CLOUD Act exposure analysis is required.

For institutions applying digital-sovereignty criteria, both the supplier and the processing are European, and the models run on the institution's own hardware.

3.4 Records of processing / DPIA

Because processing is local, an institution completing a DPIA or a record of processing activities can typically describe Airgap the way it would describe a locally-installed word processor. A DPIA usually needs categories of data, retention, location, transfers and recipients. The institution determines every one of them, not us. §5.4 documents exactly where data is stored so that retention and deletion can be described precisely.

---

4. Network behaviour: the complete list

The full, current version of this section, including the SHA-256 hash of every downloadable file and step-by-step instructions for verifying the claims with tools we do not control, is published at **https://airgap.scribewave.com/network**.

The application makes three outbound requests, two of which are optional:

Purpose	Host	When	Payload
Licence activation	AWS API Gateway, eu-west-3 (Paris)	Once, on licence key entry	Key, hashed device fingerprint, device name
Model download	`models.airgap.scribewave.com`	First run; when adding a model	None beyond the HTTP GET
Update check	`updates.airgap.scribewave.com`	Only if automatic updates are enabled — **off by default**	None beyond the HTTP GET

Notes a reviewer will care about:

- The updater is not merely defaulted to off. The update plugin is not registered at application startup unless the setting is already enabled, and enabling it requires an application restart before it can take any action.

- Licence activation has a five-second timeout and falls back to accepting the key offline. A blocked activation endpoint does not prevent use.
- After activation and model download, the application is fully functional with networking disabled. This is the single test we recommend an evaluator run first.

4.1 Optional remote AI providers, disclosed explicitly

Airgap's AI features (summaries, notes, and library chat) run by default on a model bundled with the application and executed on the device. A user may instead configure an external provider (an OpenAI-compatible endpoint, Anthropic, or Google Gemini) by supplying their own API key.

This is off by default, requires deliberate configuration, and raises a consent dialog naming the provider before anything is sent. We document it here rather than leaving it to be discovered, because an institution that prohibits such endpoints needs to know the pathway exists. In a default installation it is closed.

---

5. Security

5.1 Code signing

- **macOS** builds are signed with an Apple Developer ID certificate and notarized by Apple. The release pipeline contains a hard gate that fails the build if the bundle is ad-hoc signed or unnotarized, so an unsigned build cannot reach users.
- **Windows** installers are signed via Azure Trusted Signing.
- **Updates**, when enabled, are cryptographically signed; a package failing signature verification is refused.

5.2 Integrity of downloaded models

Every model file has its SHA-256 hash compiled into the application binary and verified after download. This holds whether the file comes from our CDN or from an institution's internal mirror (§6.3). Neither a mirror operator nor Scribewave can substitute different bytes without the installation failing. The full hash table is published at https://airgap.scribewave.com/network.

5.3 Telemetry

There is none. The application embeds no analytics, crash-reporting or product-event SDK. This is verifiable by the means described in §4.

5.4 Data at rest

All application data is held in a single directory:

- **macOS** — `~/Library/Application Support/com.scribewave.offline`
- **Windows** — `%APPDATA%\com.scribewave.offline`

This contains the library database, downloaded models, settings, the licence key, and any recordings the application itself created. Media imported from elsewhere on the disk is referenced in place and not copied. Deleting the directory removes all application data; there is no remote copy to request the deletion of.

Encryption at rest is provided by the operating system's full-disk encryption (FileVault, BitLocker), which is the institution's existing control rather than a separate mechanism introduced by this application.

5.5 Permissions requested

The application requests only the operating-system permissions its features need: microphone and system-audio capture for meeting recording and dictation, and macOS Accessibility permission if the user enables auto-paste for dictation. Features whose permissions are declined remain unavailable; the rest of the application is unaffected.

---

6. Deployment

6.1 System requirements

	Supported	Not supported
macOS	13 (Ventura) or later, Apple Silicon	Intel Macs
Windows	10 and 11, x64	ARM builds
Linux	—	No build

Performance depends materially on the machine. Airgap detects the hardware at first run, selects a transcription model sized to it, and states the expected speed before any work begins. Machines with under 8 GB of RAM run a smaller model with correspondingly lower accuracy; see §8.

6.2 Installation

Both a Windows NSIS installer (`.exe`) and a Windows Installer package (`.msi`) are published for each release, alongside the macOS bundle. The MSI is the appropriate artefact for managed deployment through Group Policy, SCCM or Intune, and supports the standard Windows Installer unattended-installation switches. Institutions planning a fleet rollout should validate the exact command in their own environment during the pilot; we will support that validation directly.

6.3 Networks that block external downloads

The single largest deployment obstacle on managed campus networks is the one-time model download, which is several hundred megabytes to a few gigabytes depending on the models selected. Three mitigations exist:

1. **Named errors.** Download failures report the actual cause (a DNS or firewall block, TLS interception, or an HTTP 403 from a content filter) rather than a generic failure, so IT can act on them.
2. **Resumable downloads.** An interrupted transfer resumes rather than restarting.
3. **Internal mirroring.** Host the model files on any internal HTTP server, preserving the filenames, and point the application at it with the environment variable `SCRIBEWAVE_MODEL_BASE_URL=https://mirror.your-institution.edu/scribewave`. Hash verification continues to run against the compiled-in values. For a fleet deployment this is the recommended configuration: it removes the external dependency entirely and makes the installation reproducible.

Institutions preferring to allow the external origin need to permit `models.airgap.scribewave.com` only.

---

7. Model inventory

Airgap runs published, openly available research models. Nothing in the processing path is a proprietary black box, and every file is listed with its hash at https://airgap.scribewave.com/network.

Model	Function
Whisper (small / medium / large-v3-turbo, quantised)	Speech recognition, 99 languages
Sortformer v2 streaming diarizer	Speaker diarization — default engine
pyannote segmentation 3.0 + WeSpeaker ResNet34	Speaker diarization — fallback engine
Silero VAD v5	Voice activity detection
GLiNER multi-PII	Personal-data detection for redaction
BGE-M3	Multilingual embeddings for semantic search
Qwen3 4B Instruct (quantised)	On-device summaries, notes and library chat — optional download

Which transcription model is used is decided by the hardware: machines with 12 GB of RAM or more run large-v3-turbo, 8–12 GB run medium, and below 8 GB run small.

7.1 Measured accuracy

On machines with 12 GB of RAM or more, Airgap runs large-v3-turbo. Its published word error rate is **4.4% in English and 5.9% in Dutch** on FLEURS, measured by the model's authors (https://github.com/openai/whisper/discussions/2363). Scribewave online, running larger models on GPUs, reaches 2.3%.

Turbo is a cost-performance choice, not a shortcut forced by the device. The largest Whisper models score a little better and run several times slower. Turbo is where that curve flattens, which is what makes unattended transcription on a normal laptop practical at all.

Three qualifications belong with those numbers, and we would rather state them than have an evaluation find them:

1. **FLEURS is clean read speech.** The same model scores 10.6% in English on Common Voice, which is crowd-sourced audio with accents and consumer microphones. A recorded interview sits between the two, and nearer the second one when people talk over each other.
2. **A word error rate says nothing about who spoke.** Our own diarization scores **7% diarization error rate**, and **cpWER is 27%** on our multi-speaker test set. cpWER counts a word correct only when the speaker label is also right, which is why it is far larger than any figure in the table above. It is the number that matches what you read in a real transcript.
3. **Accuracy follows the model your hardware runs**, not the application. The medium and small tiers are meaningfully weaker; see §6.1 and §8.

Speaker labelling is the weaker half of this product, and closing that gap is what an institutional licence funds. Making local transcription and diarization more accurate is the direction of the work, stated as a direction and not as a delivery date (§9).

7.2 Languages

Airgap transcribes the 99 languages Whisper supports, with automatic detection and an optional language hint per file. Published error rates exist for 42 of them, on both benchmarks, and the full per-language table is at **https://airgap.scribewave.com/languages**.

Languages not written with spaces between words (Chinese, Japanese, Korean, Thai) are scored by character error rate rather than word error rate; the two are not comparable and the table marks which is which.

---

8. Known limits, stated plainly

An evaluation that surfaces these later is worse for both of us than one that starts with them.

- **Word accuracy is close to the best available, not equal to it.** Scribewave online reaches 2.3% against turbo's 4.4% in English (§7.1). Users with acoustically hard recordings should be routed there.
- **Diarization is the weaker half.** 7% diarization error rate on our own test set, against 4.4% word error rate for the transcription (§7.1). Overlapping speech and large numbers of speakers remain hard, and this is where we are putting the work.
- **The on-device language model is small.** Summaries and library chat are useful, and are not comparable to a frontier hosted model.
- **Results depend on the user's hardware.** A machine below 8 GB of RAM runs a smaller transcription model and will produce noticeably weaker transcripts than a well-specified one. For a campus deployment this is the variable most worth checking against the actual estate.
- **No Linux build, and no Intel Mac build.**
- **No central administration.** Installations are independent; there is no console for managing users or auditing usage.
- **No third-party benchmark of our own build.** The FLEURS figures in §7.1 are published for the model, not for this application; our own evaluations are internal and not independently verified. Both can be shared on request, labelled as such.

---

9. Licensing and commercial terms

Individual licence — €79 one time. Up to three devices. Twelve months of updates included. After that window the installed version continues to work indefinitely; renewal is required only to install newer releases.

Institutional site licence — annual, priced by institution size, with perpetual fallback on the same principle: the versions installed during a licensed year keep working after it ends. A licence covers unlimited installations within the licensed scope, with no per-hour or per-user metering. Current indicative bands are €3,000–€5,000 per year for a single faculty or a smaller institution and €6,000–€10,000 for a large university; the exact figure depends on the population covered and is confirmed in writing before any commitment.

Many institutions find the lower band sits below the threshold that triggers a formal tender. Combined with the absence of a data processing agreement, that is usually the shortest path to approval.

What the annual fee buys. The software as installed is yours regardless. The fee funds continuous evaluation of the open speech-model field and the delivery of improved models into supported releases, plus support during the licensed year. See https://airgap.scribewave.com/mission.

10. Support

- Email support for named institutional contacts, with a defined response commitment agreed in the licence.
- Twenty-three user-facing support documents ship with the product and are indexed into the in-application assistant, so end users can self-serve without contacting either us or institutional IT.
- Direct involvement in pilot deployment, including validation of silent installation and internal model mirroring.

11. Accessibility

Not yet assessed. No formal conformance evaluation against EN 301 549 or WCAG 2.1 AA has been carried out for either the application or the website, and no conformance statement exists.

We would rather record that accurately than claim a level we have not measured. Institutions for which a conformance statement is a procurement requirement should raise it early; it is a solvable problem but not one we can answer on the day.

12. Business continuity

Scribewave is a small company, and an institution is right to weigh that. The relevant mitigation is structural rather than reassuring:

- **Perpetual fallback.** The installed version continues to function offline indefinitely. It does not phone home to stay alive, and there is no licence server whose disappearance disables it.
- **No data hostage.** Since no institutional data is held by us, there is nothing to retrieve in an exit scenario. Exports to open formats (TXT, DOCX, CSV, SRT, VTT) are available at any time.
- **Open model formats.** The models are publicly available research models in standard formats, not proprietary artefacts that vanish with the vendor.
- **Source review.** Where an institution's security policy requires source access as a condition of approval, we make source available for review under NDA to site-licence customers.

13. Evaluation and pilot process

The path we recommend, and which we will run at our own cost:

1. **Free trial.** Any staff member installs the 7-day trial. No procurement involvement, no data leaves the institution, nothing to sign.
2. **Time-boxed pilot.** An agreed group, a named end date, and a 30-minute feedback call already in the calendar. We provide extended evaluation licences for this period at no charge. Seven days is not enough time for an institutional assessment and we do not pretend otherwise.
3. **Security and data-protection review.** This document plus https://airgap.scribewave.com/network is intended to be sufficient. Where a specific questionnaire is required, send it; we complete it.
4. **Deployment validation.** Silent installation and internal model mirroring tested against the real environment during the pilot.
5. **Written quotation** the same day the scope is agreed.

Contact — hello@scribewave.com

14. Supplier details

Legal entity	Quantum Noodle BV
Trading name	Scribewave
Enterprise / VAT number	BE 1003.622.970
Registered office	Hallaardreef 5, 2580 Beerzel, Belgium
Jurisdiction	Belgium (EU/EEA)
Contact	hello@scribewave.com

A note for your finance office: the product is sold as *Scribewave Airgap*, and invoices are issued by **Quantum Noodle BV**. The two names refer to the same supplier. Please register the legal entity, not the product name, in your supplier system. A mismatch discovered at purchase-order stage is a common and entirely avoidable delay.

VAT is charged according to the applicable rules; institutions with a valid EU VAT number are invoiced under the reverse-charge mechanism where it applies. Send us your VAT number and any supplier-onboarding form with the quotation request and we will complete both before the purchase order is raised.