TranscribIA — Dossier conformité avant-vente

Version du 6 mai 2026 — à destination des DSI / RSSI / DPO

1. Présentation et positionnement

TranscribIA est une plateforme SaaS de transcription audio intelligente destinée aux organisations qui manipulent des contenus sensibles : DSI, RSSI et DPO de collectivités locales, hôpitaux, cabinets juridiques et ETI françaises. Le service combine transcription automatisée, diarisation des locuteurs et génération de texte assistée par LLM, avec des quotas configurables par organisation et par utilisateur.

La stack applicative repose sur Laravel 12 et PHP 8.3 pour le backend, Sanctum avec Bearer tokens pour l'authentification API, React 19 avec TypeScript et Vite pour l'interface, PostgreSQL 16 en architecture multi-tenant schema-per-org, Redis 7 pour le cache, les sessions et la queue, Laravel Reverb pour le WebSocket, Nginx 1.27 en reverse proxy et Docker Compose avec 8 services.

Le positionnement de TranscribIA est volontairement factuel : proposer une solution française, hébergée en France pour le serveur principal, avec stockage et sauvegardes dans l'Union européenne. L'argument de confiance repose sur des mesures vérifiables : cloisonnement par organisation, audit trail, sauvegardes chiffrées, 2FA TOTP, contrôle des rôles, antivirus sur les uploads et notification d'incident selon le cadre RGPD.

2. Hébergement et souveraineté des données

TranscribIA est hébergée sur un serveur unique Hostinger KVS situé dans le datacenter de Strasbourg, en France. Les données applicatives sont stockées sur disque NVMe local chiffré. L'architecture déclarée limite l'hébergement, le stockage principal et les sauvegardes aux environnements situés dans l'Union européenne, sans transfert hors UE pour ces fonctions.

Les sauvegardes offsite sont chiffrées avant externalisation avec age, utilisant X25519 et ChaCha20-Poly1305, puis synchronisées via rclone vers Backblaze B2 ou Scaleway en UE. Le chiffrement est donc appliqué avant sortie du serveur principal.

La plateforme n'est pas présentée comme une certification de cloud souverain auditée. Elle documente précisément les composants en place : serveur principal en France, données exclusivement UE, sauvegardes chiffrées offsite UE, sous-traitants techniques listés, et traitement IA encadré par des services identifiés. Cette transparence permet aux DSI, RSSI et DPO d'évaluer les risques sur des faits techniques concrets.

3. Sécurité technique

La sécurité applicative de TranscribIA repose sur plusieurs couches. Les accès API utilisent Laravel Sanctum avec Bearer tokens, dont les tokens sont hashés en SHA-256 en base de données. Les mots

de passe sont hashés avec bcrypt cost 12. Les connexions sont servies en HTTPS partout, avec certificats Let's Encrypt via Nginx Proxy Manager, et TLS 1.2 ou supérieur obligatoire.

L'authentification multi-facteur est disponible via TOTP conforme RFC 6238, compatible avec Google Authenticator, Authy, Microsoft Authenticator et 1Password. L'enrôlement 2FA est configurable par rôle. Les secrets 2FA sont protégés par un cast Eloquent Laravel utilisant AES-256-CBC. Les actions destructives ou sensibles — suppression, restauration ou modification de politique 2FA — nécessitent un step-up password.

Les autorisations s'appuient sur un RBAC à 5 niveaux : super_admin, agent, manager, team_lead et user. Les endpoints sensibles bénéficient du rate limiting Laravel. CORS est strict et CSP est active. Chaque upload est scanné par un sidecar ClamAV 1.4 avant écriture disque, avec limite de scan à 2 Go, signatures mises à jour via freshclam, rejet hard des fichiers infectés, log et alerte. L'isolation multi-tenant repose sur un schéma PostgreSQL distinct par organisation.

4. RGPD et protection des données personnelles

TranscribIA est conçue pour être utilisée par des clients responsables de traitement. La base légale du traitement est déterminée par chaque organisation cliente selon son contexte métier et son registre RGPD. TranscribIA intervient comme sous-traitant technique pour l'hébergement, la transcription, la diarisation, la génération assistée et la conservation paramétrable des transcriptions.

Les sous-traitants déclarés sont Hostinger pour l'hébergement en France, Backblaze B2 ou Scaleway pour les sauvegardes offsite en UE, OpenRouter comme passerelle LLM vers Anthropic/OpenAI avec no-train flag activé, et un serveur de transcription externe sur GPU dédié. Ces sous-traitants doivent être repris dans le DPA contractuel et dans la documentation RGPD remise au client.

Les droits RGPD sont couverts. L'accès et la rectification sont disponibles par édition libre des transcriptions. La portabilité, sous forme d'export complet par tenant signé en SHA-256, est disponible en self-service depuis Réglages → Portabilité (article 20 RGPD). Le droit à l'effacement (article 17) est disponible par demande certifiée depuis la même page : la requête est tracée dans audit_logs et déclenche un workflow de purge dans le délai légal avec émission d'un certificat de suppression. Les durées de conservation sont paramétrables par organisation. La traçabilité est assurée par une table activity_logs par tenant et par public.audit_logs pour les événements plateforme, exportables en CSV ou JSON pour les besoins d'audit. Le contact pour les questions de conformité est olivier@ambassadia.com.

5. Sauvegardes et continuité d'activité

TranscribIA dispose de sauvegardes quotidiennes chiffrées offsite. Le chiffrement est réalisé avec age, basé sur X25519 et ChaCha20-Poly1305, avant synchronisation par rclone vers Backblaze B2 ou Scaleway en UE. La politique de rétention prévoit 90 jours pour les sauvegardes complètes et 30 jours pour les sauvegardes par organisation.

L'objectif de RPO est de 24 heures : en cas d'incident majeur, la perte de données cible ne doit pas dépasser la dernière sauvegarde quotidienne disponible. L'objectif de RTO est de 4 heures : le service vise une restauration opérationnelle dans ce délai cible. Ces valeurs sont des objectifs d'exploitation, utiles pour cadrer les attentes contractuelles et les plans de continuité internes des clients.

Un DR drill a été réalisé sur l'organisation Test. La restauration a été vérifiée bit pour bit à l'aide des scripts backup-offsite.sh et restore-from-offsite.sh. Avant toute restauration, une safety net est appliquée : un snapshot automatique de l'état courant est réalisé afin de limiter le risque d'écrasement irréversible. Les sauvegardes et restaurations sont granulaires par organisation, ce qui permet une reprise ciblée sans restaurer l'ensemble de la plateforme lorsque le périmètre de l'incident est limité.

6. Sous-traitants techniques

TranscribIA maintient une liste explicite de sous-traitants techniques afin de faciliter l'analyse DSI, RSSI et DPO.

• Hostinger — Strasbourg, France — Serveur unique KVS, hébergement principal — À intégrer au DPA client comme hébergeur. • Backblaze B2 — UE — Sauvegarde offsite chiffrée — À intégrer au DPA client comme stockage backup. • Scaleway — UE — Alternative de sauvegarde offsite chiffrée — À intégrer au DPA client comme stockage backup. • OpenRouter — Service LLM externe — Passerelle vers Anthropic/OpenAI, no-train flag activé — À intégrer au DPA client comme sous-traitant IA. • Serveur de transcription externe — GPU dédié, hébergement UE — Transcription automatique des fichiers audio — À intégrer au DPA client comme sous-traitant transcription.

Les sauvegardes sont chiffrées avant externalisation. Les traitements LLM transitent par OpenRouter avec no-train flag activé. Le serveur de transcription externe est un GPU dédié dont les modèles sont mis à jour régulièrement. Cette présentation permet au client de vérifier les rôles, les flux et les garanties applicables avant contractualisation.

7. Gestion des incidents et notifications

La gestion des incidents repose sur la détection technique, la traçabilité et un processus de notification RGPD. Les événements applicatifs sont enregistrés dans un audit trail tenant via une table activity_logs propre à chaque organisation. Les événements plateforme sont conservés dans public.audit_logs. Cette séparation permet de distinguer l'activité propre à un client des événements transverses d'administration.

Les uploads font l'objet d'un contrôle antivirus systématique par ClamAV 1.4 avant écriture disque. Les fichiers infectés sont rejetés, journalisés et déclenchent une alerte. Les limites de scan sont définies à 2 Go et les signatures sont maintenues via freshclam. Le sidecar ClamAV est configuré en fail-open avec logs : si le scanner est indisponible, l'événement est tracé. Ce comportement doit être pris en compte dans l'analyse de risque client.

En cas de violation de données personnelles, le processus opérationnel passe par olivier@ambassadia.com. L'objectif de notification à la CNIL est de 72 heures lorsque la notification est requise par le RGPD. Les journaux d'audit, les traces applicatives, le RBAC, le step-up password sur actions destructives et les snapshots pré-restauration contribuent à qualifier l'incident, limiter son périmètre et documenter les décisions prises.

8. Identification des intervenants — diarisation et corrélation vocale (encadrement RGPD)

TranscribIA met en œuvre deux traitements distincts liés aux intervenants, d'encadrement RGPD différent.

(1) Diarisation positionnelle automatique. À l'arrivée du fichier audio, le serveur de transcription exécute une diarisation par pyannote.audio qui produit des étiquettes positionnelles anonymes (Speaker 1, Speaker 2, …) indiquant que deux blocs de la même session appartiennent au même locuteur, sans préciser qui. Aucun gabarit vocal n'est conservé au-delà de la durée de traitement du job ; les embeddings en mémoire vive sont libérés à la fin du traitement. Ce traitement positionnel, sans gabarit conservé, est non biométrique au sens de l'article 9 du RGPD (position de la CNIL, lignes directrices sur la biométrie 2019, mise à jour 2023).

(2) Attribution manuelle et aides au remplissage. L'opérateur autorisé du tenant ouvre la session transcrite et, pour chaque Speaker N, saisit un nom humain (Marie Dupont, M. le Maire, Rapporteur, …). Cette opération est strictement manuelle. Des aides au remplissage accélèrent la saisie sans jamais analyser la voix : détection de mentions de prénoms par traitement NLP du texte, autocomplétion à partir de la liste de membres connus du tenant, historique des display_name déjà saisis.

(3) Corrélation vocale éphémère (« voice correlation ») — traitement optionnel. Depuis mai 2026, TranscribIA propose une fonctionnalité désactivée par défaut permettant, au sein d'un même tenant, une comparaison vocale éphémère entre sessions afin de proposer un rapprochement entre intervenants. Ce traitement implique le calcul temporaire de caractéristiques vocales en mémoire vive, sans constitution de gabarit biométrique persistant ni d'empreinte vocale réutilisable conservée. Il ne peut être activé que par le responsable de traitement (organisation cliente), est strictement opt-in, et suppose la signature préalable d'un avenant DPA dédié ainsi que la réalisation d'une analyse d'impact relative à la protection des données (AIPD, article 35 RGPD). Les modalités techniques, les garanties de non-persistance et le périmètre exact sont décrits dans la politique de confidentialité et dans le DPA (Annexe 1 §A1.6 et Annexe 3 §A3.9).

Qualification juridique. La diarisation positionnelle sans gabarit conservé n'est pas un traitement biométrique au sens de l'article 9. La corrélation vocale éphémère porte sur des caractéristiques vocales ; bien qu'elle ne constitue pas un gabarit biométrique persistant, elle est traitée par prudence comme un traitement à risque au titre de l'article 35 et encadrée par AIPD, avenant DPA dédié, activation opt-in et transparence. Les bases légales ordinaires de l'article 6 restent déterminées par chaque tenant client en tant que responsable de traitement. La note technique d'analyse RGPD de l'identification des intervenants est disponible sur demande (format PDF/DOCX) et ses éléments-clés sont annexés au DPA. TranscribIA s'engage à ne jamais introduire de traitement biométrique automatique persistant (reconnaissance vocale avec base de voiceprints) sans une mise à jour explicite du DPA soumise à acceptation préalable du tenant, avec un préavis minimum de 60 jours.

9. Engagement de transparence et portabilité

TranscribIA met en avant une approche de transparence utile en avant-vente : les mesures en place, les dépendances, les limites et les travaux en cours sont explicitement distingués. Une DPIA est

disponible sur demande afin de faciliter l'analyse d'impact menée par les clients responsables de traitement, notamment dans les secteurs public, santé, juridique et ETI.

La réversibilité est traitée à deux niveaux. Les sauvegardes et restaurations sont granulaires par organisation grâce à l'architecture PostgreSQL schema-per-organization. L'export complet client par tenant est disponible en self-service depuis Réglages → Portabilité : un .tar.gz signé en SHA-256 contenant l'intégralité du schéma de l'organisation (sessions, audio, transcriptions, IA, votes, présences, dictionnaire, journal d'activité, utilisateurs). La suppression certifiée (article 17) est disponible depuis la même page : la requête est tracée, exécutée dans le délai légal, et fait l'objet d'un certificat de suppression remis au client.

TranscribIA n'est pas audité SOC 2 à ce stade. Un audit est prévu en 2026. Cette information est volontairement mentionnée pour éviter toute ambiguïté dans les dossiers RSSI. Les contrôles déjà en place restent vérifiables : TLS 1.2+, 2FA TOTP, RBAC, audit logs, isolation tenant par schéma, sauvegardes chiffrées offsite, DR drill validé, antivirus sur uploads et notification CNIL sous 72 heures en cas de violation notifiable.

10. Directive NIS2 — articulation des mesures techniques de TranscribIA

La directive (UE) 2022/2555, dite NIS2, impose aux entités essentielles et importantes — dont les administrations publiques centrales, régionales et locales — des obligations de gestion des risques cybersécurité, de traitement des incidents et de sécurisation de la chaîne d'approvisionnement. Elle est transposée en droit national et ses seuils de désignation relèvent du cadre de chaque État membre ; TranscribIA n'est pas une certification NIS2 — une telle certification de produit n'existe pas à ce titre — mais un outil dont les mesures techniques vérifiables peuvent contribuer à la mise en conformité de l'entité cliente, sous sa responsabilité. Sur le plan de la gouvernance et du contrôle d'accès, la plateforme applique un RBAC à cinq niveaux (super_admin, agent, manager, team_lead, user), une authentification 2FA TOTP conforme RFC 6238 configurable par rôle, un step-up password sur les actions destructives, des tokens API Laravel Sanctum hashés en SHA-256 et des mots de passe en bcrypt cost 12. Le chiffrement couvre le transit en TLS 1.2+ obligatoire, les données au repos sur disque NVMe chiffré et base PostgreSQL, les secrets 2FA protégés en AES-256-CBC et les sauvegardes chiffrées avec age (X25519 + ChaCha20-Poly1305) avant externalisation. La traçabilité est assurée par une table activity_logs propre à chaque tenant et par public.audit_logs pour les événements plateforme, exportables en CSV ou JSON.

Le traitement des incidents s'appuie sur la détection (sidecar ClamAV 1.4 sur les uploads avec signatures freshclam, surveillance active 24/7 de la queue, du disque, des certificats et des dépendances, alertes Slack/email et digest quotidien) et sur l'audit trail pour qualifier, endiguer et documenter un incident ; un processus de notification est défini dans la logique NIS2 d'alerte précoce, de notification et de rapport final, en cohérence avec l'objectif de notification à la CNIL sous 72 heures applicable en cas de violation de données personnelles notifiable. La continuité d'activité repose sur des sauvegardes quotidiennes chiffrées offsite en UE (rétention 90 jours pour les complètes et 30 jours par organisation), des objectifs d'exploitation de RPO 24h et RTO 4h, un DR drill validé bit pour bit et une restauration granulaire par organisation précédée d'un snapshot de sécurité anti-écrasement. La chaîne d'approvisionnement est documentée et reprise au DPA : hébergeur Hostinger en France, sauvegardes Backblaze B2 ou Scaleway en UE, passerelle LLM OpenRouter avec no-train flag activé, serveur de transcription sur GPU dédié en UE. En mode on-premise, la maîtrise de l'inventaire, du durcissement et de l'exploitation des contrôles incombe à

l'entité cliente ; en SaaS souverain, ces contrôles sont exploités et documentés par AMBASSADIA SARL. Le contact conformité est olivier@ambassadia.com.

11. Mesures techniques et organisationnelles de sécurité (Art. 32 RGPD)

Conformément à l'article 32 du RGPD, TranscribIA met en œuvre des mesures techniques et organisationnelles appropriées pour assurer un niveau de sécurité adapté au risque. Pseudonymisation et limitation d'accès : isolation multi-tenant par schéma PostgreSQL dédié à chaque organisation, RBAC à cinq niveaux (super_admin, agent, manager, team_lead, user), authentification 2FA TOTP RFC 6238 configurable par rôle, step-up password sur les actions destructives, tokens API Laravel Sanctum hashés en SHA-256. Confidentialité : chiffrement en transit TLS 1.2+ obligatoire (Let's Encrypt via Nginx Proxy Manager), chiffrement au repos sur disque NVMe et base PostgreSQL, secrets 2FA protégés en AES-256-CBC, URLs signées pour les ressources. Intégrité : contrôle antivirus ClamAV 1.4 systématique sur les uploads (signatures freshclam, rejet des fichiers infectés), journaux d'audit (activity_logs par tenant et public.audit_logs plateforme, exportables CSV/JSON), CSP active et CORS strict. Disponibilité et résilience : sauvegardes quotidiennes chiffrées offsite en UE avec age (X25519 + ChaCha20-Poly1305), rétention 90 jours (complètes) et 30 jours (par organisation), RPO 24h, RTO 4h, restauration granulaire par organisation précédée d'un snapshot anti-écrasement, DR drill validé, surveillance active 24/7 (queue, disque, certificats, dépendances) avec alertes. Évaluation et réévaluation régulières des risques, tests de sécurité, et processus de notification de violation à la CNIL sous 72 heures. En mode on-premise, ces mesures sont mises en œuvre sur l'infrastructure du client sous sa responsabilité opérationnelle ; en mode SaaS souverain, elles sont exploitées et documentées par AMBASSADIA SARL.