ZILLIONLABS · CETERIS-PARIBUS PRO® DOCGEN

Guide pratique auditeur

Moroccan Auditor How-To Guide

FR + EN · 6 tâches pratiques / 6 practical tasks

v1.0 · 01 November 2026

Auditeurs · DPO · Responsables conformité

Auditors · DPO · Compliance officers

Ce guide est fourni à titre informatif. Tout document généré par IA doit être révisé par un professionnel qualifié avant signature. / This guide is provided for informational purposes. Any AI-generated document must be reviewed by a qualified professional before signing.

FRANÇAIS — Guide pratique auditeur

01 — Charger votre corpus documentaire

Le corpus est l'ensemble des documents sur lesquels le système base toutes ses réponses et tous les documents générés. Aucun document ingéré = aucune génération ancrée. Commencez toujours ici.

Flux de chargement

Vos documents → Console Ingestion Gateway → Tumart Vector v3.5 → DocGen Génération

Étapes

1. Ouvrez la console principale (/) et repérez le widget Sovereign Ingestion Gateway dans le panneau central.
2. Déposez vos fichiers (.txt, .md, .json, .pdf, .docx). Les PDF scannés passent automatiquement par la pipeline vision OCR page par page.
3. Cliquez « Ingest to Pro® ». Un toast vert « N chunks ingérés » confirme le succès.
4. Vérifiez GET /api/status pour l'inventaire complet : filename, chunk_count, ingested_at, source_url.

■ *Formats acceptés : .txt · .md · .markdown · .json · .pdf · .docx · .doc · images jpg/png/tiff (vision OCR)*

---

02 — Ingérer des PDF et des URL

Le Sovereign Ingestion Gateway prend en charge les PDF (y compris scannés via OCR vision) et les pages web. Ces méthodes enrichissent votre corpus sans téléchargement manuel.

Méthode A — Upload de PDF

Glissez-déposez votre fichier .pdf dans le widget Ingestion. Le système détecte automatiquement si le PDF est textuel (extraction directe) ou scanné (vision OCR). Les PDF scannés nécessitent que VISION_MODEL et VISION_BASE_URL soient configurés.

Méthode B — Ingestion par URL

Collez une ou plusieurs URLs (une par ligne) dans le widget « Add from URL ». Activez « 1 level deep » pour suivre les liens du même domaine (jusqu'à 20 sous-pages). Chaque page est extraite par trafilatura et ingérée comme fichier .md.

■ *Limites : timeout 15 s par page · max 2 MB par page · max 20 sous-pages par URL racine · http/https uniquement.*

---

03 — Générer un document de conformité

Le DocGen génère des documents juridiques marocains structurés à partir de votre corpus. Chaque document est ancré dans les textes ingérés — jamais inventé. La génération se fait en mode formulaire (FORM) ou en mode question libre (CHAT).

Flux de génération

Sélection du doc type → Formulaire ou question → Tumart retrieval top-8 → CP Pro® NDJSON stream → Document DOCX/PDF

Choisir le bon type de document par phase

Phase 1 — Incorporation SA : Certificat Négatif OMPIC, Statuts SA, Bulletins de souscription, DSV, Lettres d'acceptation, Mandat CAC, Bail commercial/domiciliation.

Phase 2 — Conformité CFC : Demande d'accréditation CFC, Charte de conformité CFC, Convention de domiciliation, Rapport de conformité réglementaire, Déclaration ANRT opérateur.

Phase 3 — IPO / Marché financier : Note d'information AMMC, Rapport de due diligence financière, Rapport d'évaluation indépendante, Mémorandum d'information investisseurs, Convention d'introduction en bourse, Rapport de conformité AMMC post-introduction.

Étapes

1. Naviguez vers /docgen/ via le lien « DocGen » dans la console principale.
2. Cliquez sur le type de document souhaité dans le panneau gauche. Le badge MODE (FORM ou CHAT) indique le mode de saisie.
3. Mode FORM : remplissez les champs obligatoires (marqués *). Plus les champs sont précis, plus le document est ancré dans votre corpus.
4. Mode CHAT : décrivez en langage naturel le document souhaité avec les parties concernées, la date et le contexte spécifique.
5. Cliquez « Générer le document ». Le texte s'affiche progressivement en streaming.

■ *Plus les champs sont précis, plus le document généré est ancré dans votre corpus. Évitez les descriptions vagues.*

---

04 — Interpréter les citations et le score de confiance

Chaque réponse du système est ancrable sur des sources précises. Les citations et scores sont le mécanisme d'audit central.

Indice de confiance juridique (Legal Confidence Index)

Le score (0–100 %) est calculé comme suit : références citées / (références citées + clauses non couvertes). Un score ≥ 80 % indique que la quasi-totalité des clauses sont étayées par le corpus ingéré. Un score < 60 % doit déclencher une vérification approfondie ou une réingestion de documents manquants.

Lire les citations dans le texte

Le modèle est contraint par son prompt système à citer le(s) nom(s) de fichier(s) pour chaque affirmation, sous la forme [filename.md :: chunk N]. Si aucune citation n'apparaît sur une affirmation,

traitez-la avec prudence — elle peut être une inférence du modèle non directement fondée sur le corpus.

Score cosinus

Valeur entre 0 et 2. Un score < 0,4 indique une très forte pertinence. Un score > 0,9 indique une correspondance faible — le document généré risque de manquer de fondement dans le corpus.

> ■ *Vérifiez toujours les citations contre les documents sources originaux avant de signer ou de transmettre un document généré.*

---

05 — Exporter et partager les documents générés

Le DocGen exporte les documents générés au format DOCX (Word) ou PDF directement depuis l'interface.

Étapes d'export

1. Attendez la fin du streaming (le curseur clignotant disparaît). Relisez le document dans le panneau principal avant d'exporter.
2. Cliquez « ■ .docx (Word) » ou « ■ .pdf » dans la section Télécharger. Le fichier est téléchargé immédiatement.
3. Le format DOCX est recommandé pour la révision et la signature électronique — vous pourrez modifier le texte dans Word après l'export.
4. Tout document généré par IA doit être révisé par un professionnel qualifié avant signature ou transmission officielle.
5. Transmettez le document exporté via les canaux sécurisés de votre organisation. Ne partagez pas les documents confidentiels via des services cloud non approuvés.

> ■ *Le document généré est un projet (draft). Sa valeur juridique dépend de la révision et de la signature d'un professionnel habilité.*

---

06 — Que faire quand le système dit « Pas dans le corpus »

Le moteur Ceteris-Paribus Pro® est configuré pour déclarer explicitement quand il ne peut pas répondre à partir des documents ingérés. Ce comportement est une fonctionnalité — il protège contre les hallucinations.

Réponse type du système

Le modèle génère **[CLAUSE NON COUVERTE PAR LE CORPUS]** lorsqu'une obligation légale ne peut pas être étayée par le corpus fourni. Un score de confiance bas accompagne généralement cette situation.

Actions correctives

1. Vérifiez le corpus existant : consultez la liste des documents ingérés (panneau Corpus ou GET /api/status). Vérifiez que le document attendu a bien été ingéré avec un nombre de chunks non nul.

2. Ingérez le document manquant : retournez au Sovereign Ingestion Gateway et ingérez le fichier ou l'URL manquant. Pour les textes réglementaires marocains, les sources officielles sont : BO (Bulletin Officiel), portail CNDP, site SGG.ma.

3. Reformulez la question avec des termes plus proches du vocabulaire du texte source (termes juridiques exacts, numéros d'articles, acronymes officiels).

4. Vérifiez la qualité de l'ingestion : PDF scanné sans VISION_MODEL configuré ? Langue du document incompatible avec le modèle d'embedding ? Contenu dans des tableaux ou images (traitement limité) ?

5. Ne jamais accepter une réponse sans citation : une réponse sans [filename :: chunk N] doit être traitée avec la même prudence qu'une réponse « pas dans le corpus ».

■ *Si aucune de ces solutions ne fonctionne, le document ne contient peut-être pas l'information demandée. Consultez les sources primaires directement (Bulletin Officiel, portail CNDP, SGG.ma, AMMC).*

ENGLISH — Auditor How-To Guide

01 — Load your document corpus

The corpus is the set of documents on which the system bases all its answers and all generated documents. No ingested document = no grounded generation. Always start here.

Loading flow

Your documents → Console Ingestion Gateway → Tumart Vector v3.5 → DocGen Generation

Steps

1. Open the main console (/) and locate the Sovereign Ingestion Gateway widget in the central panel.
2. Drop your files (.txt, .md, .json, .pdf, .docx). Scanned PDFs automatically go through the page-by-page vision OCR pipeline.
3. Click 'Ingest to Pro®'. A green toast 'N chunks ingested' confirms success.
4. Check GET /api/status for the full inventory: filename, chunk_count, ingested_at, source_url.

■ *Accepted formats: .txt · .md · .markdown · .json · .pdf · .docx · .doc · images jpg/png/tiff (vision OCR)*

---

02 — Ingest PDFs and URLs

The Sovereign Ingestion Gateway supports PDFs (including scanned via vision OCR) and web pages. Both methods enrich your corpus without manual downloads.

Method A — PDF Upload

Drag and drop your .pdf file into the Ingestion widget. The system automatically detects whether the PDF is text-based (direct extraction) or scanned (vision OCR). Scanned PDFs require VISION_MODEL and VISION_BASE_URL to be configured.

Method B — URL Ingestion

Paste one or more URLs (one per line) into the 'Add from URL' widget. Enable '1 level deep' to follow same-domain links (up to 20 sub-pages). Each page is extracted by trafilatura and ingested as a .md file.

■ *Limits: 15s timeout per page · max 2 MB per page · max 20 sub-pages per root URL · http/https only.*

---

03 — Generate a compliance document

DocGen generates structured Moroccan legal documents from your corpus. Each document is anchored in the ingested texts — never invented. Generation uses either form mode (FORM) or free-text mode (CHAT).

Generation flow

Select doc type → Form or free text → Tumart retrieval top-8 → CP Pro® NDJSON stream → DOCX/PDF Document

Choosing the right document type by phase

Phase 1 — SA Incorporation: OMPIC Negative Certificate, SA Articles of Association, Subscription bulletins, DSV, Acceptance letters, CAC mandate, Commercial lease/domiciliation.

Phase 2 — CFC Compliance: CFC accreditation application, CFC compliance charter, Domiciliation agreement, Regulatory compliance report, ANRT operator declaration.

Phase 3 — IPO / Capital markets: AMMC information note, Financial due diligence report, Independent valuation report, Investor information memorandum, Stock exchange introduction agreement, Post-IPO AMMC compliance report.

Steps

1. Navigate to /docgen/ via the 'DocGen' link in the main console.
2. Click the desired document type in the left panel. The MODE badge (FORM or CHAT) indicates the input mode.
3. FORM mode: fill in the required fields (marked *). The more precise the fields, the more the document is grounded in your corpus.
4. CHAT mode: describe the desired document in natural language with the parties involved, date, and specific context.
5. Click 'Generate document'. The text appears progressively via streaming.

■ *The more precise the fields, the more the generated document is grounded in your corpus. Avoid vague descriptions.*

---

04 — Interpreting citations and confidence scores

Every system response can be anchored to precise sources. Citations and scores are the core audit mechanism.

Legal Confidence Index

The score (0–100%) is calculated as: cited references / (cited references + uncovered clauses). A score >= 80% indicates nearly all clauses are grounded in the ingested corpus. A score < 60% should trigger a thorough review or re-ingestion of missing documents.

Reading inline citations

The model is constrained by its system prompt to cite source filename(s) for each claim, in the form [filename.md :: chunk N]. If no citation appears on a statement, treat it with caution — it may be a model inference not directly grounded in the corpus.

Cosine score

Value between 0 and 2. A score < 0.4 indicates very high relevance. A score > 0.9 indicates weak matching — the generated document may lack grounding in the corpus.

> ■ *Always verify citations against the original source documents before signing or transmitting a generated document.*

---

05 — Export and share generated documents

DocGen exports generated documents in DOCX (Word) or PDF format directly from the interface.

Export steps

1. Wait for streaming to complete (the blinking cursor disappears). Review the document in the main panel before exporting.
2. Click '■ .docx (Word)' or '■ .pdf' in the Download section. The file downloads immediately.
3. DOCX is recommended for review and electronic signature — you can edit the text in Word after export.
4. Any AI-generated document must be reviewed by a qualified professional before signing or official transmission.
5. Transmit the exported document via your organisation's secure channels. Do not share confidential documents via non-approved cloud services.

> ■ *The generated document is a draft. Its legal value depends on review and signature by an authorised professional.*

---

06 — What to do when the system says 'Not in corpus'

The Ceteris-Paribus Pro® engine is configured to explicitly declare when it cannot answer from the ingested documents. This behaviour is a feature — it protects against hallucinations.

Typical system response

The model generates `[CLAUSE NON COUVERTE PAR LE CORPUS]` when a legal obligation cannot be supported by the provided corpus. A low confidence score generally accompanies this situation.

Corrective actions

1. Check the existing corpus: consult the list of ingested documents (Corpus panel or GET /api/status). Verify that the expected document has been ingested with a non-zero chunk count.
2. Ingest the missing document: return to the Sovereign Ingestion Gateway and ingest the missing file or URL. For Moroccan regulatory texts, official sources are: BO (Official Gazette), CNDP portal, SGG.ma site.
3. Rephrase the question using terms closer to the source text vocabulary (exact legal terms, article numbers, official acronyms).
4. Check ingestion quality: scanned PDF without VISION_MODEL configured? Document language incompatible with the embedding model? Content in tables or images (limited processing)?

5. Never accept an unsourced response: a response without [filename :: chunk N] citations should be treated with the same caution as a 'not in corpus' response.

■ *If none of these solutions work, the document may not contain the requested information. Consult primary sources directly (Official Gazette, CNDP portal, SGG.ma, AMMC).*