**TextLab — Pitch Deck**

A lightweight, slide-by-slide outline. Copy to slides and add screenshots.

---

**Slide 1 — Title**

\*\*TextLab\*\*: Turn PDFs into insights for evaluation teams

• Bulk download → Extract → Consolidate → Analyze → Summarize → Chat

• Private, local, fast

**Slide 2 — Problem**

• Evaluation reports are long, inconsistent, and multilingual

• Analysts spend days copy‑pasting sections and counting terms

• Topic discovery and cross‑country comparisons are ad‑hoc

**Slide 3 — Solution**

• End‑to‑end workflow on your laptop

• User‑entered headings drive consistent extraction

• Keyword analytics (exact + semantic), topic modeling, and smart summaries

• Private Q&A chat grounded only in your data

**Slide 4 — Why Now**

• Proliferation of PDFs; need faster evidence synthesis

• Mature local NLP (SBERT/BM25/BERTopic) works offline

• Growing demand for transparent, auditable analysis pipelines

**Slide 5 — Product Tour (1/2)**

• Bulk PDF Downloader (CSV/Excel links) with robust filename and direct‑download fixes

• PDF→TXT with best‑available extractor (PyMuPDF/pdfplumber/PyPDF2)

• Consolidate by user headings; map year/country/region from your sheet

**Slide 6 — Product Tour (2/2)**

• Keyword Analyzer with validation snippets and Year/Region charts

• Document Topics (BERTopic/NMF) + semantic quality + top docs

• Summaries (overall/findings/keywords) + private evaluation chat

**Slide 7 — Differentiation**

• 100% local: no data leaves your machine

• Headings are user‑defined per corpus (fits UNICEF/COAR variability)

• Safety clamps to avoid ML parameter crashes; graceful fallbacks

**Slide 8 — Impact**

• 5–10× faster synthesis vs manual extraction

• More recall with semantic matching; fewer missed findings

• Repeatable outputs (CSV downloads) for audit and re‑use

**Slide 9 — Architecture (high‑level)**

• Streamlit UI; session state for consolidated dataframe

• PDF backends: PyMuPDF → pdfplumber → PyPDF2

• NLP: BM25/TF‑IDF + SBERT; BERTopic/NMF/LDA

• Optional: Argos Translate for offline language unification

**Slide 10 — Use Cases**

• COAR cross‑country synthesis

• Education/Teacher evaluations — capacity, retention, training

• Humanitarian sitreps trend analysis across regions/years

**Slide 11 — Roadmap**

• OCR fallback (Tesseract) for scanned PDFs

• Presets for heading packs; project templates

• Model registry for embeddings; GPU acceleration

• Export to DOCX/PowerPoint

**Slide 12 — Call to Action**

• Pilot with 100–500 PDFs

• Identify 3–5 standard heading packs

• Define reporting templates for summaries and keyword dashboards

---

**Speaker Notes (quick hits)**

• Emphasize local/privacy; “no data leaves your machine.”

• Show validation snippets to build trust in counts.

• Use semantic search live to answer a country‑specific question.

• End with a CSV export—stakeholders like takeaways.