**TextLab — Pitch Deck**

Slide 1 — Title

TextLab: Turn PDFs into insights for evaluation teams

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

Slide 2 — Problem

• Long multilingual reports; inconsistent headings

• Manual copy/paste and term counts

• Ad-hoc cross-country comparisons

Slide 3 — Solution

• End-to-end workflow on your laptop

• User-entered headings for robust extraction

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

• Private Q&A chat grounded only in your data

Slide 4 — Why Now

• Proliferation of PDFs

• Mature local NLP (SBERT/BERTopic) offline

• Demand for transparent, auditable pipelines

Slide 5 — Product Tour (1/2)

• Bulk PDF Downloader

• PDF→TXT extraction

• Consolidate by headings; map year/country/region

Slide 6 — Product Tour (2/2)

• Keyword Analyzer with validation snippets + charts

• Document Topics + semantic quality + top docs

• Summaries + private chat

Slide 7 — Differentiation

• 100% local, privacy-first

• User-defined headings; UNICEF/COAR-friendly

• Safety clamps; graceful fallbacks

Slide 8 — Impact

• 5–10× faster synthesis

• Better recall via semantic matching

• Reusable CSV outputs

Slide 9 — Architecture

• Streamlit UI, session state

• PDF: PyMuPDF/pdfplumber/PyPDF2

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

• Optional: Argos Translate

Slide 10 — Use Cases

• COAR cross-country synthesis

• Teacher evaluations: capacity, retention, training

• Humanitarian sitreps trends

Slide 11 — Roadmap

• OCR fallback; heading presets; model registry; GPU; export to DOCX/PPTX

Slide 12 — Call to Action

• Pilot with 100–500 PDFs; define 3–5 heading packs; set reporting templates.