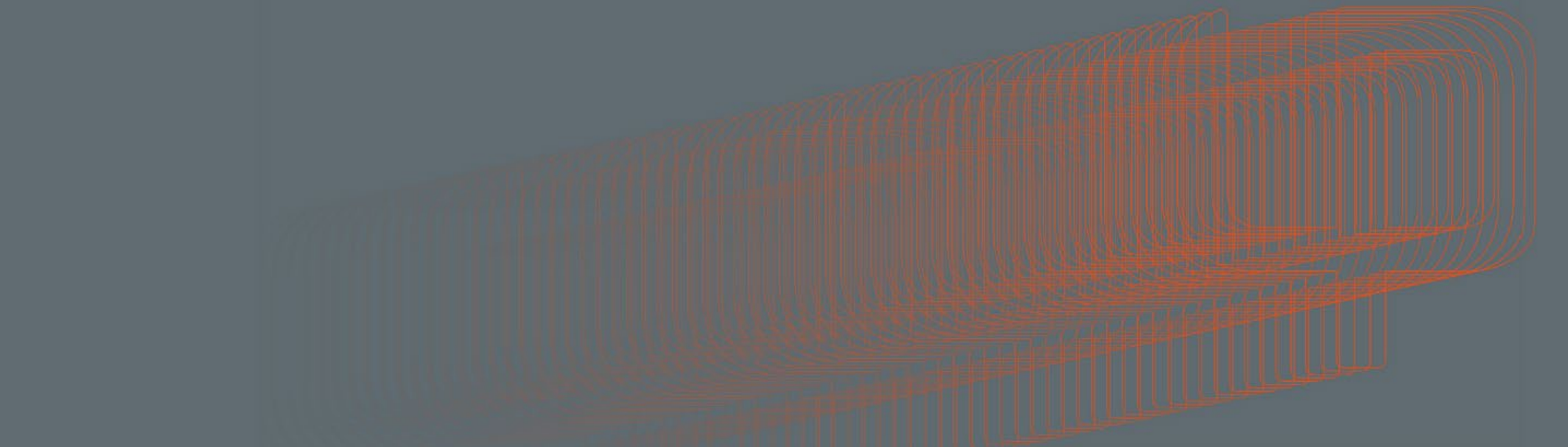
uphios
we empower your business

Transforming future.

Together.

SignaQuery:
Signature Detection & Query Engine



About Me

Alberto Gregorio

Technical Director | Architect

MSc Computer Science

alberto.gregorio@phios.group

Ruggell, Liechtenstein

15+ Jahre Entwicklung von Softwarelösungen

Enterprise-Softwarearchitektur

Entwicklung robuster verteilter Systeme, skalierbarer APIs und sauberer Kerndomänen.

Integration moderner Tech Stacks & KI

Förderung cloudnativer Innovation, Data Intelligence und missionskritisches Engineering.

Technische Strategie & Lieferexzellenz

Die Vision des Kunden in sauberen Code übersetzen, hohe Sicherheit und nachhaltige Lieferung.

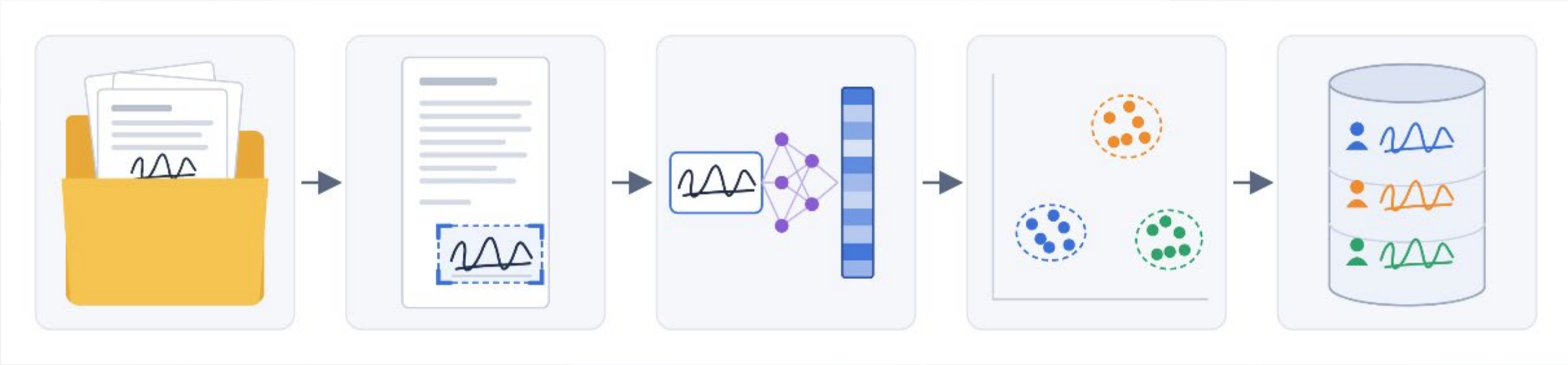
PHIOS AG

- Founded in Ruggell, Liechtenstein, in 2007
- A team of 40 consultants and developers
- Locations in Ruggell, Valencia, and Vienna
- IT Engineering
- Development of high-quality digital solutions

Problemformulierung

- Die hohe Anzahl juristischer Dokumente erschwert die Zuordnung und Gruppierung nach beteiligten Personen.
- Für juristische Dokumente **reicht OCR nicht aus**, da handschriftliche Texte oder Unterschriften nicht verlässlich erkannt werden.
- Aus rechtlichen Gründen (Datenschutz & Vertraulichkeit) muss der gesamte Prozess **on-premises** betrieben werden.

Ingestion Pipeline-Schritte

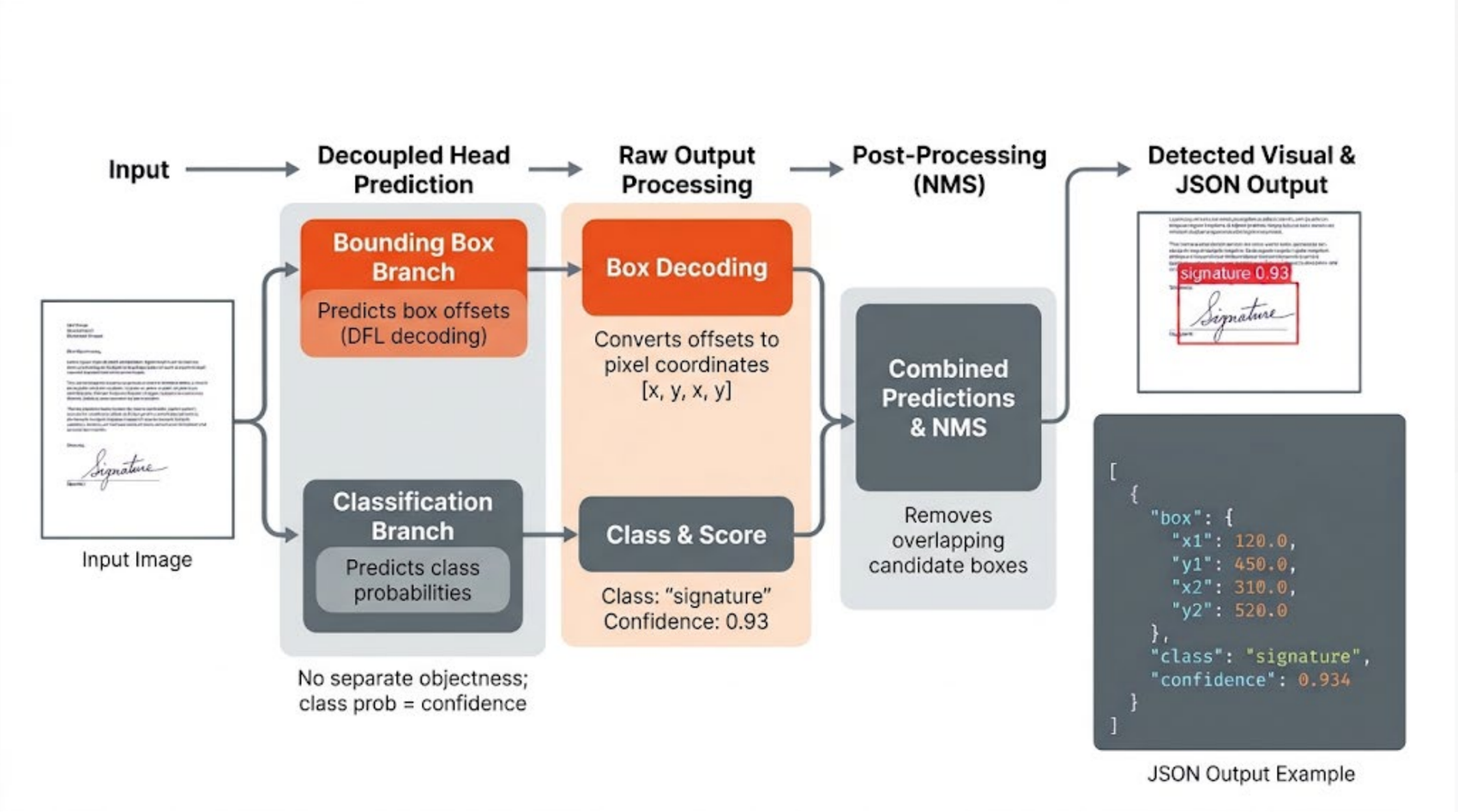


Folder Detection Embedding Clustering Database

SignaQuery: Technischer Überblick

- **SignaQuery** ist eine Windows-Desktop-Anwendung, die von der **PHIOS AG** für die Dokumentenabrufung basierend auf visuellen Funktionen wie Signaturen und Handschrift entwickelt wurde.
- Die Desktop-App sendet Dokumentenseiten- und Signaturzuschnitt-Bilder an Laufzeit-APIs.
- Die APIs erkennen Signaturen, bereinigen die Daten, berechnen Embeddings und clustern die resultierenden Vektoren.
- **SignaQuery** unterstützt einen ähnlichkeitsbasierten Abruf ohne Textgenerierungsschritt.

YOLOv8 als Erkennung

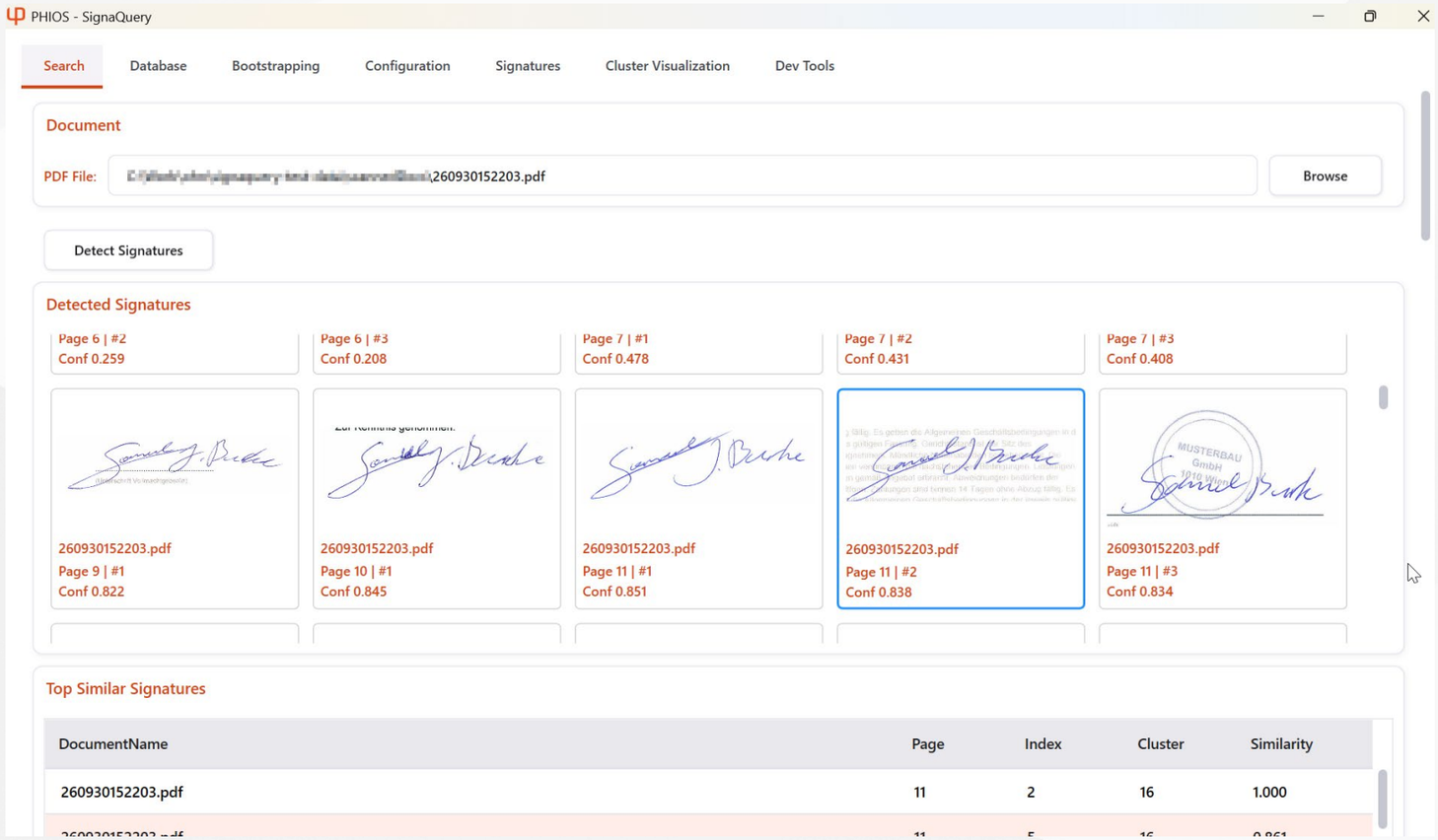


SignaQuery: Cleanup (1)

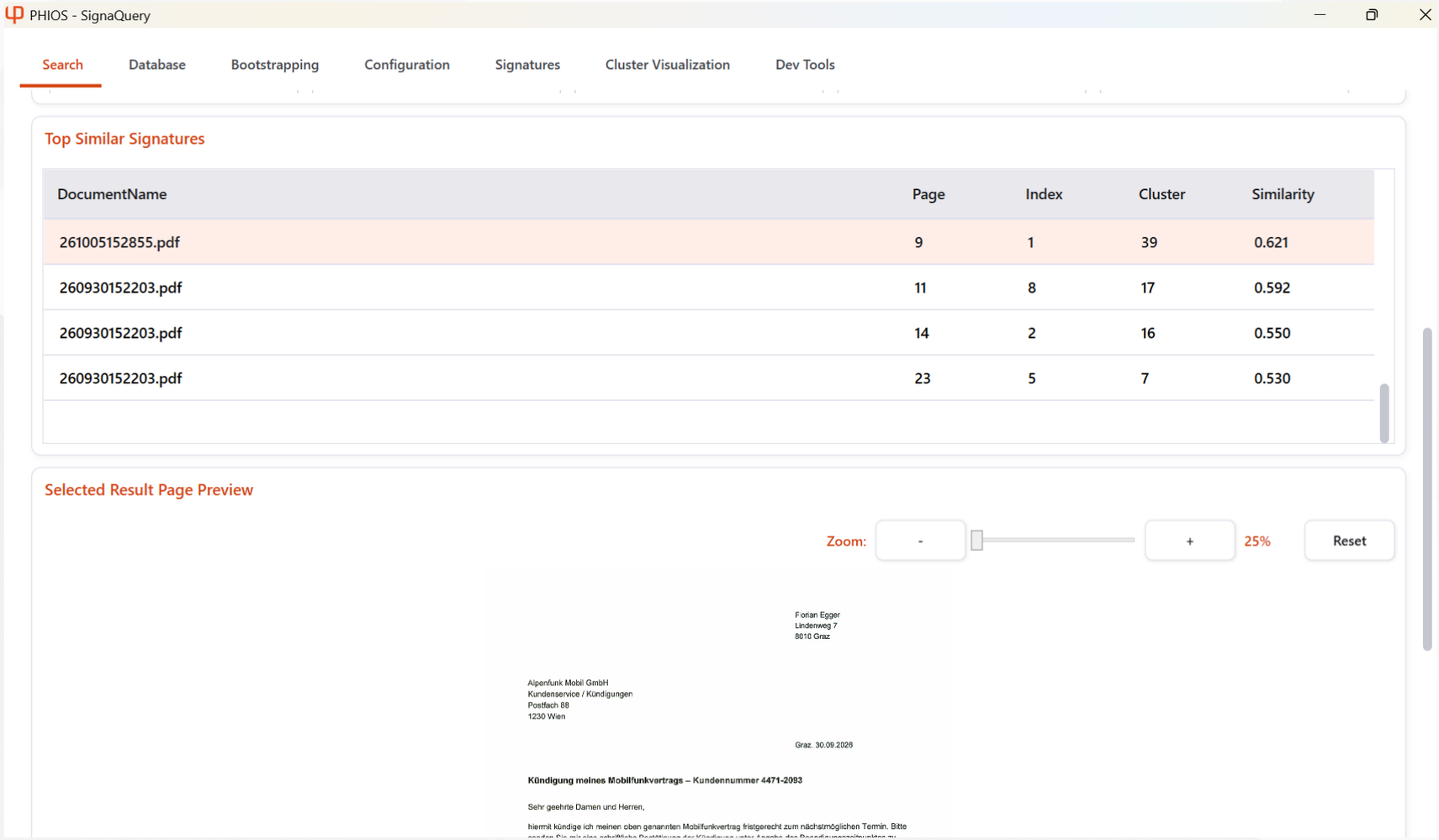
- Erforderlich aufgrund der sehr hohen Anzahl (>80 %) an **Singleton-Clustern**
- Für jeden erkannten Signaturschnitt (Crop):
  - Tinte erkennen und farbige Tinte (z.B. blauen Stift) von neutraler Tinte unterscheiden. Farbige Tinte für den nächsten Schritt zwischenspeichern.
  - Lange, dünne horizontale und vertikale Linien der neutralen Tinte entfernen.
  - Fallback auf die vorherige Tintenmaske nutzen, falls die Filterung zu aggressiv war.
  - Bereinigtes Bild an das Embedding-Modell übergeben.

SignaQuery: Cleanup (2)

SignaQuery: Similarity Search (1)



SignaQuery: Similarity Search (2)



Gelernte Lektionen

- **Ready-to-use Open-Source-Modelle** zu finden, ist einfacher denn je – allerdings decken sie spezifische Anwendungsfälle selten exakt ab und ein Finetuning ist meist kostenintensiv.
- Es gibt eine wachsende Zahl von Unternehmen und Einzelpersonen, die an **Local-First-Lösungen** interessiert sind, da sie ihre Dokumente nicht in die Big Clouds der Hyperscaler hochladen wollen oder aus rechtlichen Gründen nicht hochladen dürfen.
- **Spec-Driven-Anwendungen** bringen dich nur bis zu einem gewissen Punkt. Der letzte Test, die Bewertung und das Debugging erfordern immer noch eine ordentliche Portion Man-In-The-Loop-Erfahrung.

Future-proof through digitalization.

With PHIOS AG