**Resume Intelligence**

Hiring for requestions entails screening large volumes of resumes to extract relevant information and rank candidates effectively. In this scenario, the resume dataset is stored as a compiled PDF.

Here’s the challenge outlined for you and the dataset is attached.

The main challenges are:

1. Efficiently use GenAI/Traditional AI to extract meaningful fields from resumes, capturing more than just the basics
2. **Uncover Non-Trivial Insights :** Go beyond surface-level data to derive actionable and unique insights about candidates
3. Establishing robust evaluation criteria to validate extraction accuracy & compute the extraction accuracy measures
4. Build a scalable, modular Python solution to ensure reusability and adaptability

**Modular Code Structure**:

* **resume\_parser/**:
  + extractor.py: Functions to extract fields from raw text
  + insights.py: Logic to compute non-trivial insights
  + evaluator.py: Metrics calculation and validation routines
  + utils.py: Shared utilities (e.g., text cleaning, PDF parsing helpers)
* **tests/**:
  + Unit tests for each module to ensure correctness and maintainability
  + Test data with known ground truth for reliable evaluation
* **examples/**:
  + Sample scripts demonstrating how to use the package
* **README.md**:
  + Documentation on installation, usage, and customization