### **Entity Extraction from Loan Agreement PDFs**

We are working on a project to automatically extract key information from Loan Agreement PDFs using AI. To ensure the solution meets your exact needs, please provide your input on the following points.

### **1. Trigger**

* How will the extraction pipeline be triggered?  
  **Client's Response:***Please provide details here...*

### **2. Content**

**2.1 PDF Source**

* From where will we source the PDFs? Will they be from an email attachment or another storage location?
* Please provide the necessary details to access the source (e.g., connection details, directory structure, how to identify the correct files).  
  **Client's Response:***Please provide details here...*

**2.2 Input Schema**

* Could you please provide a fixed format for the input schema?
* Will the input be a simple JSON with keys and data types, or will it be in a different format that requires pre-processing? Please share a sample of the structured schema.  
  **Client's Response:***Please provide details here...*

### **3. Evaluation**

* Should we use the existing evaluation pipeline or create a new one for this specific use case?
* If we proceed with the existing internal pipeline, we will require these three inputs:
  1. One Schema file (JSON)
  2. Sample PDFs
  3. Ground truth data for each sample PDF (in CSV format)

**Client's Response:***Please provide details here...*

### **4. Output Processing**

* Once the final entities are extracted (as JSON or CSV), what is the downstream application or workflow that will use these results?
* Does our team need to handle this integration, or is there an existing process in place?  
  **Client's Response:***Please provide details here...*