**📄 Project Report: Collateral Description Agent**

**Student:** Li Kailang(FDQAOR)

**🛠️ System Components**

| **Module** | **Description** |
| --- | --- |
| BLIP-2 | Vision-language model used to generate English image captions |
| EasyOCR | Extracts visible text in images (supports English, Hungarian) |
| Qwen-32B-AWQ via API | Language model used to generate final markdown reports |
| FastAPI | Serves the system via a Dockerized REST API |
| Markdown Writer | Automatically saves descriptions to .md files in the output/ directory |

**⚙️ Workflow**

1. **Input:** A folder of 20–50 images per asset is uploaded to the local endpoint /generate.
2. **Processing:**
   * BLIP-2 generates natural image captions
   * EasyOCR extracts multilingual visible text
   * These results are combined and used as prompt input
   * Qwen API generates detailed markdown description including:
     + Identification
     + Condition Assessment
     + Documented IDs and accessories
     + Estimated value (optional)

**✅ Results**

* Successfully processed multiple asset folders
* Description accuracy significantly improved when both vision and OCR are used
* Markdown format allows easy integration with other reporting or loan systems
* Response time for a 30-image asset is ~20–30 seconds on a CPU-based system