

EXTERN + OUTAMATION  
EXTERNSHIP

---

# ADVANCED AI-POWERED DOCUMENT INSIGHT AND DATA EXTRACTION PROJECT

---

SPRINT 1 STATUS UPDATE

by Lashawn Fofung, MBA, PMP, CSM, CSPO

LF

## SKILLS

- Python
- AI/ML
- Data Scrapping
- Prompt Engineering
- LLMs
- Text Extraction
- Document Intelligence
- Python Automation
- Optical Character Recognition (OCR)
- Retrieval-Augmented Generation (RAG)

Helping businesses improve, automate, and architect critical tasks (such as document handling, customer service, and data management) by saving thousands of hours and boosting accuracy.



# SPRINTS



- 01 AI Document Intelligence
- 02 Python & Google Colab
- 03 Python Data Extraction
- 04 Optimizing OCR
- 05 Implementing RAG
- 06 Optimizing RAG Pipelines
- 07 Blob Processing & Classification
- 08 Interactive Chatbot
- 09 AI-Powered Document Automation Platform

---

PROJECT TIMELINE

---

10 WEEKS

# AI Document Intelligence

Extracting Key Insights from Mortgage Files

# Background: 10-Week Project

The Outamation Advanced AI-Powered Document Insights and Data Extraction Externship is a comprehensive 10-week program. Its core goal and purpose is to master the required skills to create an AI-Powered Documentation Automation Platform. The program focuses on bridging foundational technologies (ML, LLMs, NLP, Computer Vision) with high-value industry applications, such as solving complex, document-heavy challenges in mortgage automation, ultimately transforming slow, manual review processes into automated, efficient workflows.

A faint, semi-transparent graphic overlay on the right side of the slide. It features a network diagram with nodes and connections. Below it are two forms: a 'BANK STATEMENT' form with fields for 'NAME', 'GROSS PAY', 'HIRE DATE', 'AMT OF MORTGAGE', 'ACCOUNT NUMBER', and 'ACCOUNT RURME'; and a 'MORTGAGE APPLICATION' form with similar fields. The overall theme is digital data and automation.

# Overall Status: Sprint 1 Complete

I have successfully concluded Sprint 1 focused on Advanced AI-Powered Document Insights and Data Extraction. The primary goal of this sprint was to establish both the theoretical foundation and the practical application of AI in document automation. I progressed seamlessly from mastering core AI concepts, such as LLMs and Generative AI, to implementing industry-grade tools. I am now fully prepared to apply this end-to-end document intelligence pipeline to solve real-world business challenges.



# Key Accomplishments

My key achievements center on validating the full document-to-data workflow:

- **Integrated Tech Stack Mastery**
  - I mastered the critical steps of converting document visuals into meaningful data, combining Computer Vision (CV) and OCR for digitization with Natural Language Processing (NLP) for deep textual comprehension.
- **Hands-on Tool Execution**
  - I gained practical experience by successfully utilizing powerful platforms, including Google's Document AI, to perform high-accuracy, real-time Document Classification and Data Extraction from complex, unstructured files.
- **Business Context**
  - I applied these learnings directly to Mortgage Automation, defining how AI accelerates loan processing and minimizes human error, providing tangible business value through speed and precision.

# Key Metrics & Progress

This sprint resulted in measurable progress in both technical and soft skill development:

- **Technical Proficiency**
  - Achieved validation of my ability to convert diverse documents into structured, usable data using a top-tier tool (Google Document AI).
- **Process Improvement**
  - I identified critical inefficiencies in manual review, positioning me to help implement solutions that can shift document processing time from weeks to minutes.
- **Skill Level Up**
  - The exercise significantly improved my Attention to Detail and Critical Thinking, which are essential skills for performing crucial tasks like Anomaly Detection and fraud prevention across large, complex document packages.

---

THANK YOU

---

<https://github.com/LashawnFofung/AI-Powered-Document-Automation-Platform>