

# Codebase & Reproduction Guide

---

## Codebase Structure & Execution

### 1. Directory Overview

- `paper/`: LaTeX source files, TikZ diagrams (`overview_tikz.tex`), figures, and presentation slides.
- `code/plotting/`: Scripts generating progressive performance curves and slide assets (`generate_ppt_assets.py`).
- `code/evaluation/`: Metric computation (CER, WER, edit distance, cost modeling).
- `ppt_assets/`: Rendered publication assets (PDF/PNG curves for baseline, random, ConfBERT, ours, safeguard, oracle).
- `DocEng presentation.pptx`: Official presentation slide deck.

### 2. Key Commands

```
# Clone the repository
git clone https://github.com/Stergios-Konstantinidis/Cost-Aware-Human-LLM-Collaboration-for-post-OCR-Corrections.git
cd Cost-Aware-Human-LLM-Collaboration-for-post-OCR-Corrections

# Generate publication and presentation assets
python3 code/plotting/generate_ppt_assets.py

# Overleaf synchronization (pull before modify, push after)
```

```
git pull origin main
```

```
git push origin main
```

---

Revision #3

Created 2026-09-09 16:20:17 UTC by Stergios

Updated 2026-09-17 12:21:52 UTC by Stergios