

# **National University of Computers and Emerging Sciences FAST-NUCES**

**BSAI(A,B,C,D)**

## **Project Phase 1: Virtual Company**

### **(Learn Professionalism Technically)**

#### **Task 1: Basic Image-to-Word Converter (MVP)**

**Problem:** Converting scanned documents or images to editable Word format typically results in plain, unformatted text. Users lose essential formatting like bold/italic styles, alignment, and paragraph structure, requiring manual reconstruction.

**Objective:** Develop a desktop application that extracts text from images while preserving basic formatting. The tool will accept JPG/PNG inputs, process them using Tesseract OCR, detect formatting attributes, and generate a formatted .docx file. Sample images are attached.

**Scope:** Single-page, English-language documents with clear text. Includes image preprocessing, OCR integration, formatting detection (bold, italic, alignment, paragraphs), and Word document generation with a simple GUI (Tkinter/PyQt).

**Deliverables:** Working prototype with source code, project report, and demonstration video.

**Outcome:** A functional converter that maintains document formatting, reducing post-processing time. This foundation enables advanced features (equations, multi-language) in later phases.

**Requirement:**

**Make group of 3 students**

**Submission deadline : Wednesday 28 – 1- 2026**