Project Proposal: Automated Information Extraction from Egyptian National ID

1. Project Overview:-

This proposal outlines the development of an AI-powered system to extract and process key information from Egyptian National ID cards. The goal is to automate the extraction of structured data, such as name, national ID number, date of birth, and other relevant details, using computer vision and Natural Language Processing (NLP) techniques. This system aims to enhance efficiency, reduce manual entry errors, and integrate seamlessly with existing databases or applications.

2. Proposed Solution:-

To achieve accurate and reliable extraction, we propose the implementation of an OCR-based pipeline combined with AI-driven text processing, including:

- **Image Pre-processing**: Enhance ID card images by adjusting contrast, noise reduction, and alignment.
- **Optical Character Recognition (OCR)**: Use Tesseract OCR or Google Vision API to extract text from the ID card.
- **Named Entity Recognition (NER)**: Implement NLP models to structure extracted text into fields such as name, date of birth, and ID number.
- **Validation & Verification**: Cross-check extracted data using checksum algorithms or predefined rules to ensure accuracy.
- **API Deployment**: Develop an API interface for seamless integration with other systems.

3. Tools & Technologies:-

The following technologies and tools will be utilized:

- **Programming Language**: Python
- **OCR Engine**: Tesseract OCR, Google Vision API, or AWS Textract
- **Image Processing**: OpenCV, PIL
- **Text Processing & NER**: SpaCy, Transformers (Hugging Face), or custom-trained deep learning models

- **Database & Storage**: PostgreSQL, MongoDB, or Firebase (depending on your requirements)
- **Deployment**: FastAPI or Flask for API development
- **Cloud & Hosting**: AWS, Azure, or on-premises setup as per your preferences

4. Project Timeline:-

The project will be executed in four phases over a 6-week timeline:

Phase	Task	Duration
1	Requirement Analysis & Data Collection	1 week
2	Model Development (OCR, NER, Image Processing)	2 weeks
3	API Development & Testing	2 weeks
4	Deployment, Fine-tuning & Documentation	1 week

5. Deliverables:-

- A fully functional AI-powered ID extraction system with API support.
- Source Code & Documentation explaining the model and workflow.
- User Guide for system integration and API usage.
- Optional: Web Interface for manual verification (if required).

6. Benefits & Expected Outcomes:-

- **Automated Data Entry**: Eliminates the need for manual ID data input, reducing errors.
- **High Accuracy & Speed**: Ensures rapid and precise extraction of ID details.
- **Scalable & Secure**: The system can process a large volume of IDs while maintaining data security.
- **Seamless Integration**: Can be integrated with government systems, banks, and enterprise applications.

7. Next Steps:-

Upon approval, we will proceed with:

- Finalizing project scope and client-specific requirements.
- Setting up the development environment and acquiring sample ID images.
- Developing and optimizing the extraction pipeline.

We look forward to your feedback and are happy to discuss any modifications or additional features to align the project with your needs.