MCQ Automated Evaluator

1. Design a Standard Answer Sheet Template

- Fixed layout (e.g., name, roll number, bubbles for each question A–D)
- Clearly defined regions for questions and answers
- Printable and scannable with good contrast
- Use grid-based layouts (like OMR sheets) for consistent detection.

2. Collect and Preprocess Answer Sheets

- **Image Input** - Use scanned copies or high-res photos (ensure consistent resolution, lighting)
- **Preprocessing**

Grayscale conversion

Noise reduction (cv2.GaussianBlur)

Thresholding (cv2.adaptiveThreshold or cv2.threshold)

Skew correction (cv2.getRotationMatrix2D + cv2.warpAffine)

Resize if needed for consistency

3. Detect Answer Regions (ROIs)

- Use fixed coordinates if the layout is standard. Or dynamically detect boxes using : Contour detection (cv2.findContours)
- Shape filtering (rectangles of a certain area/aspect ratio)
- Sorting to get question-wise order
- Optional : Use template matching if alignment varies.

4. Bubble Detection (Answer Recognition)

For each question:

- Segment the answer area into 4/5 options
- Count black pixels (filled bubbles have more)
- Use threshold to decide which option is marked

5. Match with Answer Key

- Store correct answers in a dictionary: answer_key = {1: 'A', 2: 'C', ...}
- Compare student answers to key
- Assign marks (e.g., +1 correct, 0 wrong/unattempted)

6. (Optional) AI-Based Enhancements

- CNN Model for detecting marked/unmarked Label bubbles as filled/unfilled manually for a small dataset
- Train a binary classifier (e.g., with Keras/TensorFlow)
- Use pytesseract to extract text
- Preprocess text regions to improve OCR accuracy

7. Generate and Save Results

- Store scores in CSV/Excel: roll_no, name, score
- (Optional) Generate a visual feedback sheet with detected answers marked
- Use matplotlib or cv2.putText() to annotate the sheet

🧠 Tips

- Start with rule-based methods → Then add AI where needed
- Build a small labeled dataset if using DL
- Normalize image sizes/aspects to reduce variation
- Use a standard printed sheet for better accuracy
- Consider adding confidence thresholds to reject ambiguous inputs