**Step 1: Designing the OCR-Friendly Sheet**

**1. The Standardized Template Layout**

**Key Design Features:**

* **Header Section:** Include a clear space for the **FinqyID** and the **Date**.
* **Grid for Leads:** Each row in the grid will represent a single lead (customer).
* **Checkbox/Bubbling for Codes:** For Slot, Connectivity, and Disposition, the caller should **not** write the word. They will either check a box or fill in a bubble next to the appropriate code. This is the single most important change for accuracy.

**2. Developing Memorable & OCR-Friendly Codes**

**A. Slot Codes:**

Instead of writing "10 am - 11 am", we'll use a simple sequential code.

* **S1:** 10:00 AM - 11:00 AM
* **S2:** 11:00 AM - 12:00 PM
* **S3:** 12:00 PM - 1:00 PM
* **S4:** 1:00 PM - 2:00 PM
* **S5:** 2:00 PM - 3:00 PM
* **S6:** 3:00 PM - 4:00 PM
* **S7:** 4:00 PM - 5:00 PM
* **S8:** 5:00 PM - 6:00 PM

**How to Mark on the Sheet:** The caller will simply write the code (e.g., "S3") in a designated box.

**B. Connectivity & Disposition Codes:**

This is a two-part process. The disposition depends on the connectivity.

**Part I: Connectivity**

* This should be a simple choice: **Y (Yes)** or **N (No)**.

**Part II: Disposition Codes (Mnemonic - Based on Memory)**

**If Connectivity = Y (Positive Outcomes):**

* **IN** - Interested
* **NI** - Not Interested
* **CB** - Call Back
* **FU** - Follow Up
* **MI** - Want More Info
* **LB** - Language Barrier
* **CD** - Call Drop

**If Connectivity = N (Negative/No-Contact Outcomes):**

* **RG** - Ringing
* **SO** - Switch Off
* **IV** - Invalid Number
* **OS** - Out of Service
* **WN** - Wrong Number
* **BZ** - Line Busy

**How to Mark on the Sheet:** The caller writes the two-letter code in the "Disposition" box.

3. **OCR Engine**: The uploaded image is fed into an OCR engine. This engine will perform the following actions:

* **Image Pre-processing**: The system will automatically clean the image—adjusting for skew, contrast, and brightness to ensure optimal character recognition.
* **Template Recognition**: It will identify the standardized template to understand where each data field is located.
* **Data Extraction**: The OCR will read the text and handwritten information from each field (Name, Mobile No., Disposition Code, etc.).

4. **Data Validation & Structuring**: This is a critical step to ensure data quality.

* **Automated Validation**:
* **Confidence Scoring**: The OCR will assign a confidence score to each extracted field. For example, a clearly written "Y" gets a 99% score, while a messy scribble might get 40%.

5. **Human-in-the-Loop (HITL) Verification**:

* Any data with a low confidence score (e.g., below 85%) is flagged for manual review.
* A data entry operator or manager sees the original image snippet and the OCR's interpretation side-by-side in a simple interface. They can quickly correct any errors. This ensures near-perfect data accuracy.

6. Portal Pages:

**Page 1 : Upload Data**

This is the core data-entry page.

* **Elements:**
  + A large, clear "Upload Files" box with drag-and-drop functionality.
  + An "Or Browse Files" button to open a file selector.
  + **Instructions Text:**
    - "Upload Sheets (JPG, PNG, PDF) or a Standardized Excel File (.xlsx)."
    - "You can upload multiple files at once."
    - A link to "Download the Standard Excel Template".
  + **Upload Progress Bar:** Shows the status of the file uploads.

**Flow:**

1. User drags or selects their files (can be a mix of images and Excel files).
2. The system uploads the files.
3. **Backend Logic:**
   * If the file is an **Excel file**, the system validates the headers. If they match, the data is directly ingested into the database. If not, an error is shown.
   * If the file is an **image/PDF**, it is sent to the OCR processing queue.
4. A success message appears: "[Y] files were uploaded successfully. [Z] records have been added. You will be notified if any records require manual review."

**Page 2: Validate Data**

This is the Human-in-the-Loop (HITL) interface for quality control.

* **Elements:**
  + A table listing all records flagged by the OCR system for review. Columns: Caller ID, Date, Reason for Flag (e.g., "Low Confidence", "Logic Error").
  + When a record is clicked, a modal window or a new section appears.
  + **Validation Interface:**
    - **Left Side:** The cropped image of the specific row from the original sheet.
    - **Right Side:** A form with the data the OCR extracted. Fields with low confidence are highlighted in red.
    - Buttons: "Approve" and "Save & Next".

**Flow:**

1. The manager clicks on a record to review.
2. They compare the image to the extracted data.
3. They correct any errors in the form fields (e.g., change NI to IN).
4. They click "Approve". The clean data is sent to the database, and the record is removed from the queue.