Main idea to solve in both the challenges: Fintech companies get messy and unstructured data (eg: calls and documents), this type of unstructured data is easy to understand for humans but difficult for an ai model as an ai model only works with numerical data -> cant really make predictions out of a lot of messy support calls.
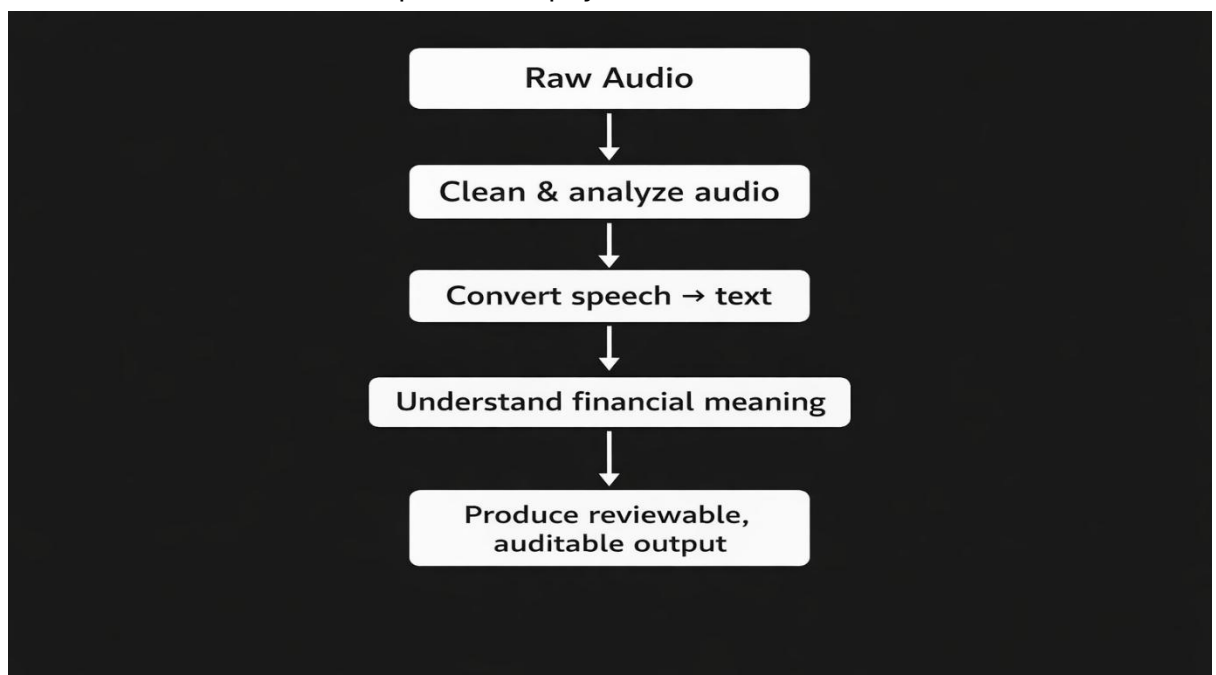
The solution we need is:

1) Input: audio calls/ scanned pdfs/ images
2) Output: Clean structured usable data (good enough to use to train a machine learning model)

Problem 1:

Banks and lenders record **millions of phone calls**:

- Customer onboarding
- KYC verification
- Payment reminders
- Complaints
- Legal consent calls
  -→ these calls are completely unstructured and can provide a lot of legal risks and unreliable stuff like -> "I promise ill pay tomorrow"

To work with the raw audio we need to understand the sound itself

1) Accept any av format and normalize them
2) Detect noise
3) Detect language

4) Speaker diarization -> who spoke when? And separation of agent vs customer
5) Flagging of untrustworthy call qualities
6) Fraud detection
   Output: clean trustworthy audio stream with the metadata

   We then need to perform financial transcription:
   Normal text to speech isn't enough cuz of financial terms so mistakes here would be a nightmare of legality

   Here the ideal output should be a time aligned transcript with a confidence rating.
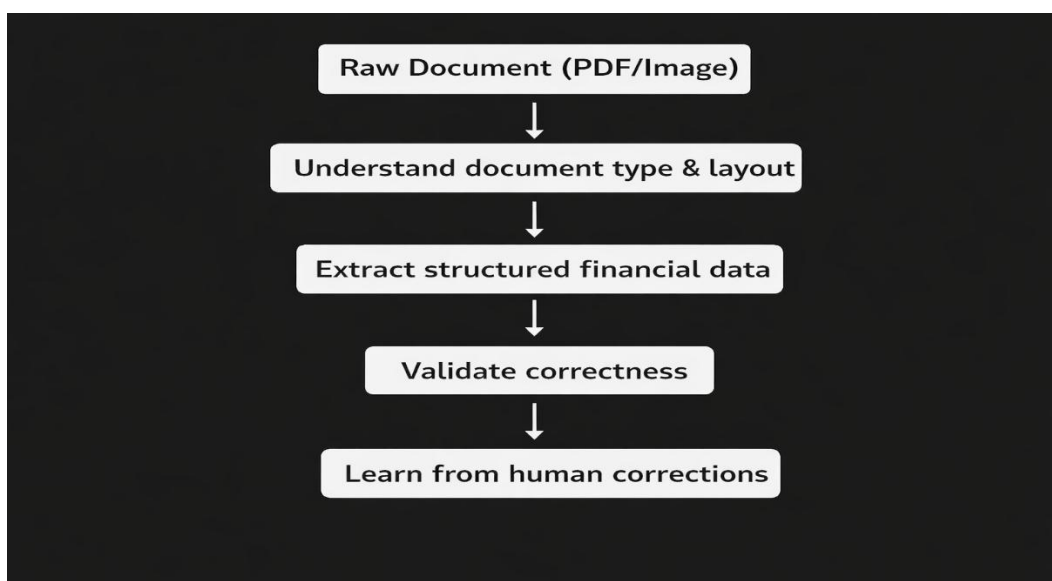
   The next step is the actual use of intelligence in this:
   1) Intent classification -> is customer agreeing or refusing or asking for time
   2) Financial entity extraction -> eg: emi amt, dates etc
   3) Obligation detection
   4) Regulatory phrasing

Problem 2:

Lenders receive hundred of document types like bank statements salary slips, invoices, loan agreements etc

The issue here is: constant change of format, manual checking isn't scalable and rule based ocr breaks

Before reading the content of the document we must first understand its layout and structure:

1) Auto document classification: bank statement vs invoice vs payslip etc etc
2) Layout detection: Tables headers handwritter notes etc
3) Multi language support
4) Image quality scoring (confidence levels here)

Now in extraction the system will read the contents in the structure and extract data and facts:

The output of this should be json type financial data or a csv

The next step is validation:

1) Cross document validation -> salary slips match bank credit , consistent employer name -> if not flag etc etc
2) Data sequencing -> no future salaries and backward transactions
   Output: Data is valid/ flagged/ needs review

   Another important thing not to miss is human in the loop -> a lot of steps taken here need to be constantly validated by a human as in financial matters a single mistake can have devastating consequences.