Prompting guidance

Building effective fields requires making strategic decisions about field types and writing well-crafted prompts that extract accurate data from documents.

Community users have access to a limited set of field types, but can typically achieve similar results using document reasoning fields with more specific prompts.

Processing mode considerations

Your project's processing mode impacts prompting effectiveness.

Agent mode uses next-generation models that excel with concise, direct instructions. These models understand context and document structure with minimal guidance.

Legacy mode uses traditional processing methods that benefit from detailed, explicit instructions. These models perform best with comprehensive context and clear formatting guidance.

Choosing field types

Select field types based on document characteristics, data structure, and performance requirements.

Field Type	Use for	Token efficiency	Source highlighting
Text extraction	Short strings, consistent formatting	Lower	Excellent
Table extraction	Structured tabular data	Medium	Full table
List extraction	Multiple items with attributes	Medium	Good
Document reasoning	Complex logic, calculations, summaries	Higher	Limited
Visual reasoning	Layout analysis, images, styling	Medium	Page-level
Derived	Values based on existing fields	High	None
Custom function	External data, complex calculations	Variable	None

Extraction vs. reasoning

Extraction fields provide predictable results and accurate source highlighting. Reasoning fields offer more flexibility but produce more variable results.

Your project's processing mode determines the optimal approach.

Agent mode

INSTABASE

Legacy mode

In legacy mode, document reasoning fields perform better for large documents over 200 pages, even for straightforward fields like names or addresses.

Document reasoning works particularly well when:

- Relevant information appears in multiple locations throughout the document.
- Contextual understanding or inference is required.
- Documents are dense or have complex formatting.

Writing effective prompts

Follow these general best practices to create clear, specific prompts.

- **Keep prompts concise** while including necessary detail.

  Policy holder's full name

  Effective date from policy summary

- **Provide context** to distinguish between similar data points or explain document structure when needed.

  Effective date from policy summary (not application date)

- **Specify formatting requirements** when basic standardization is required. For complex formatting, extract first, then clean.

  Phone number in format (XXX) XXX-XXXX

  Agent mode

  Agent mode often handles formatting instructions directly in field prompts without needing separate cleaning steps.

  Legacy mode

  Legacy mode might require additional cleaning prompts or functions for complex formatting needs.

- **For list extraction fields**, limit attributes to 10 for greatest accuracy (30 maximum).

- **For document reasoning fields**:

  - Use the **Enhance prompt** option to optimize clarity and effectiveness automatically.
  - Specify date and time if needed, otherwise the current date and time is used.
  - For chain of thought reasoning, use numbered steps when you want to see the model's logic. For a clean final result, use direct statements instead.

  Agent mode

  Reserve reasoning fields for true abstract reasoning beyond what's in the document. For complex extraction tasks, use the appropriate extraction field type instead.

INSTABASE

requires contextual understanding.

When using reasoning fields for extraction tasks, **use declarative commands** that state exactly what you want extracted.

Extract the policy holder's full name as it appears on the insurance declaration page.

- **For derived fields**:
  - Order fields so that referenced fields precede derived fields.
  - When referencing table or list extraction fields, normalize results as text using cleaning.

Extracting tables

Extract tables using extraction fields.

Legacy mode

Reasoning fields can be used for advanced filtering or refinement of existing tables, but table extraction fields typically provide more predictable results.

Follow these prompting best practices to extract tables from documents:

- **Enable table recognition** in digitization settings.

  To see recognized tables in a document, select the prediction icon in the header and enable **Show detected objects** for tables. Tables in the document are highlighted and you can use the adjacent table icon to view, copy, or download a table.

- **Identify tables by descriptive characteristics** rather than position or appearance.

  Claims Summary table containing columns for Date, Description, Amount, and Status.

- **Specify sheet name** for multi-sheet Excel files.

  Transaction table from the Account Activity or Transaction History sheet

- **Apply filtering and manipulation** through prompt instructions.
  - *Transactions with amounts greater than $1,000*
  - *Transactions for 01 April through 15 April*
  - *Transactions sorted by amount from smallest to largest*

Multipage tables might not be extracted correctly unless they have consistent headers on all pages. Source highlighting for tables indicates entire tables, not individual rows, columns, or cells.

Extracting checkboxes, signatures, and barcodes

Extract information from certain visual objects, including checkboxes, signatures, and barcodes. The type of information you can extract differs by object type.

INSTABASE

Signatures	Presence, signatory details, and dates. Extraction of signature images isn't available.
Barcodes	Presence, quantity, and encoded values.

Extract visual objects using extraction fields.

> For best results when extracting checkboxes, enable checkbox recognition in digitization settings.

For most object extraction, descriptive field names often provide adequate results:

- *Filing Status* (for grouped checkboxes)
- *Signatory name*, *Signatory title*, *Signature date*
- *Barcode value*

For standalone objects or more specific requirements, use prompts or descriptions:

- *Is the filer claiming capital gains or losses?* (standalone checkbox)
- *Extract all signatures* or *Return yes if this document is signed*
- *Return all barcode values* or *Return yes if this document contains a barcode*

Common patterns for reasoning fields

These proven prompt patterns address scenarios typical of reasoning fields. Adapt them to your specific field requirements.

> Model responses can vary even with well-structured prompts. Test patterns with your specific documents and refine as needed to achieve consistent results.

Chain-of-thought reasoning

Use numbered steps when you want to see the model's reasoning process in the result.

```
1. Locate all medication entries in the prescription history section.
2. Extract the dosage for each medication.
3. Identify any medications exceeding 500mg dosage.
4. Return a list of medications exceeding 500mg with their dosages.
```

For a clean final result without intermediate reasoning, use declarative commands instead.

```
Identify all medications in the prescription history that exceed 500mg dosa
```

Date-based selection with fallback logic

Extract values associated with the most recent date, with fallback options when dates aren't available.

```
Identify all lab results with collection dates.
Select the result with the most recent date.
```

INSTABASE

Conditional extraction

Extract data only when specific criteria are met, with clear handling of edge cases.

```
Extract the policy effective date only if:
- The date appears in the policy summary or declarations page.
- The date is within the last 2 years.
- The date format is valid (MM/DD/YYYY).
Otherwise, return "Date not found or invalid".
```

Multi-source data consolidation

Search multiple document sections in priority order and return the first valid result.

```
Determine the patient's current treatment plan:
1. Check "Treatment Plan" section for active interventions.
2. If not found, analyze "Provider Notes" for treatment recommendations.
3. Cross-reference with "Prescription History" to identify ongoing medicati
4. Synthesize findings into a summary of current treatment approach.
5. Return "No active treatment documented" if no information found.
```

Categorization

Categorize values using a predefined set of categories with clear criteria.

```
Determine the claim status as one of: [Approved/Denied/Pending/Under Review
Base the determination on:
- Explicit status statements in the document.
- Decision language (approved, rejected, requires additional information).
- Presence of approval codes or denial reasons.
If status can't be determined, return "Status unclear".
```

Extraction with categorization

Extract numeric values and automatically categorize them into meaningful ranges.

```
Extract the patient's blood pressure reading and categorize:
- If systolic under 120 and diastolic under 80, return "Normal: [value]".
- If systolic 120--139 or diastolic 80--89, return "Elevated: [value]".
- If systolic 140+ or diastolic 90+, return "High: [value]".
- If blood pressure not found, return "Reading not available".
```

Aggregation and calculation

Perform calculations across multiple values found throughout the document.

```
Calculate total claim amount:
1. Locate all line items in the itemized charges section.
2. Extract the dollar amount for each line item.
3. Sum all amounts.
4. Return total in format: $X,XXX.XX.
5. If no line items found, return "No charges listed".
```

INSTABASE

```
Determine if pre-authorization is required:
1. Search for terms: "pre-authorization", "prior authorization", "pre-appro
2. If found, return "Yes - pre-authorization required" with the specific re
3. If explicitly stated as not required, return "No - pre-authorization not
4. If unclear, return "Pre-authorization requirement not specified".
```

Was this page helpful? Yes No