

Generative Extraction — Prompting Best Practices

IXP Module 05, Part 2 — quick reference for writing effective prompts at every level.

The quality of your prompts directly determines extraction accuracy. Well-written instructions unlock the full potential of the generative model.

Project-Level Instructions

- Briefly describe the **industry** and **document type** (e.g., "This is a US payroll paystub from the HR/Finance industry")
- Mention any special characteristics of the document (multi-page, scanned, handwritten sections)
- If the document may contain multiple instances of the same data, include a **unique identifier field** in every field group

Field Group-Level Instructions

- Choose **clear, recognizable names** — use the same terminology as in the documents
- Group **related data points** together logically
- Explain how the field group contributes to the document's overall meaning
- Indicate the likely **location or section** in the document (e.g., "Found in the top-right header section")
- Treat each field group as a **table** — each field is a unique column
- Use > to create parent-child relationships (e.g., *Paystub* > *Deductions*)

Field-Level Instructions

- Start every instruction with **"Extract..."**
- Specify the exact **format and structure** of the expected output (e.g., "Extract as DD/MM/YYYY")
- Provide **example inputs and expected outputs** for complex or ambiguous fields
- Focus on **one piece of data per instruction** — don't overload a single field
- Select the appropriate **data type**: Text, Number, Date, or Boolean
- For inferred fields, describe the derivation logic (e.g., "Extract net pay — calculated as Gross Pay minus total deductions")

General Prompting Rules

Rule	Guidance
Clarity & Simplicity	Use clear, direct, unambiguous language. Avoid complex nested sentences.
Consistency	Use the same terminology across project, field group, and field instructions.
Provide Context	Give the model relevant industry and document-type context at the project level.
Iterate	Write → Test → Edit → Save. Repeat until extraction quality is satisfactory.
Avoid Negative Instructions	Instead of "Don't skip sections", say "Ensure all earnings sections are captured".

Rule	Guidance
Avoid Repetition	Repeating instructions across levels causes confusion and reduces accuracy.
Avoid Contradictions	Ensure project, field group, and field instructions do not conflict with each other.
Positive Examples	Include examples of correct responses to guide the model (few-shot).

♦ **IMPORTANT**

Exact vs Inferred Fields: Generative Extraction can extract both *exact* fields (values explicitly stated in the document) and *inferred* fields (values derived or calculated from document content). Always be explicit in your field instruction about which type each field is.

For the full best practices reference, see the official IXP User Guide:
docs.uipath.com/ixp/automation-cloud/latest/user-guide/best-practices