

NLP to Semantics

Workflow

1. User Query (Input Layer)

- The user types or speaks a query in natural language.
- Example: "Show Marketing employees earning above 50k excluding interns."
- This is the raw input that the system will understand and process.
- Queries can be **complex or ambiguous**, so the AI must handle synonyms, conditions, and context.

NLP (Natural Language Processing Layer)

- Extract **meaning, intent, and entities** from the natural language query.
- **Steps:**
 1. **Tokenization:** Break query into words or phrases.
 - ◆ "Show", "Marketing", "employees", "above 50k", "excluding interns"
 2. **Part-of-Speech Tagging:** Identify nouns, verbs, adjectives → helps GPT-4 understand roles of words.
 3. **Named Entity Recognition (NER):** Detect entities like Department = Marketing, Salary > 50k, Role ≠ Intern.
 4. **Intent Detection:** Determine the **action the user wants** → in this case, **fetch employee records**.
- Without understanding the user's intent and entities, the system cannot generate correct structured queries.

Semantic Understanding Layer

- Understand **meaning, context, and relationships** between entities.
- **Example:**
 - User may write: "List all IT staff earning more than 60k"
 - Semantic understanding allows the system to know:
 - ◆ "IT staff" = "Information Technology employees"
 - ◆ "more than 60k" = Salary > 60000
- GPT-4 is used here because it can understand:
 - Synonyms
 - Multi-condition queries
 - Context

Query Generation Layer (Semantic → Structured Query)

- Convert the **semantic understanding** into a **structured query** (SQL / OData) that the database or Azure AI Search can execute.
- **Process:**
 1. GPT-4 receives **intent + entities** from NLP/semantic layer.
 2. GPT-4 generates **executable query**
`SELECT * FROM Employee WHERE Department='Marketing'`

AND Salary>50000

AND Role<>'Intern';