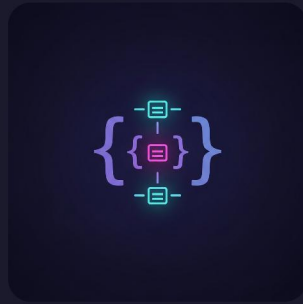


# Anatomy of an evaluator

How Arize runs that prompt against your traces



## Data fields

### WHAT IT IS

The prompt names the span attributes to read as placeholders, which the evaluator fills with real trace data at run time.

### WHY IT MATTERS

Binds the judge to the actual input, output, and context of each span rather than letting it guess.



## Eval model

### WHAT IT IS

The model that actually runs the judgement on each span or trace, scoring it against your prompt.

### WHY IT MATTERS

The flexibility lives here: swap in a cheaper or stronger judge without rewriting a single line of the prompt.



## Output

### WHAT IT IS

Each run returns a label, an optional score, and an explanation, landing on the span as structured eval attributes.

### WHY IT MATTERS

Constrained, machine-parseable results are what let you aggregate, filter, alert, and track quality over time.