arize

# Three jobs evals do

Evaluation is more than a test you run once before shipping

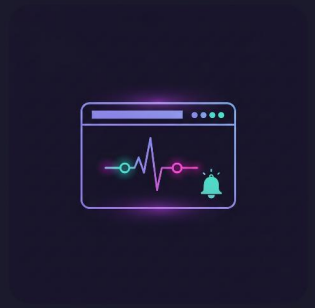
## Pre-release testing

WHAT IT MEANS

Run evals on a golden dataset before you ship, the way you run unit tests.

WHY IT MATTERS

Catches regressions while they are still cheap — before a single user sees them.



## Production monitoring

WHAT IT MEANS

Score live traffic continuously once the app is out in the real world.

WHY IT MATTERS

Real inputs drift from your test set. Monitoring is how you find out before users do.

## Inline guardrails

WHAT IT MEANS

Evals that run inside the request path, checking each response as it is produced.

WHY IT MATTERS

Blocks or corrects a bad or unsafe output in real time, before it reaches the user.

BEFORE RELEASE

IN PRODUCTION