Evaluating AI Agents

Cost · Frequency · Output Quality · Hallucinations

When designing or choosing an AI agent, there are **four core factors** you should always evaluate.There is **no perfect agent** — only **well-balanced trade-offs**.

1. Cost

What it means: How expensive the agent is to run.

Key points:

- Each AI call has a cost
- More steps = higher cost
- Larger models = higher cost

Example:A daily quote agent that runs once per day is cheap.A customer support agent running constantly is more expensive.

Rule of thumb:

> Cost becomes a problem when agents scale.

2. Frequency

What it means: How often the agent runs.

Examples:

- Daily → once per day
- Event-based → only when triggered
- Real-time → constantly

Why it matters:

- Higher frequency increases cost
- More executions = more chances for failure

Rule of thumb:

> Frequency decides how often you pay for the agent.

3. Output Quality

What it means: How useful, readable, and consistent the output is.

High-quality output is:

- Relevant
- Well-formatted
- On-topic
- Usable without manual editing

What improves quality:

- Clear prompts
- Guardrails
- Narrow scope

Rule of thumb:

> Good output is output you don’t need to fix.

4. Hallucinations

What it means: When an AI confidently produces incorrect or made-up information.

Common examples:

- Fake facts
- Invented quotes
- Incorrect explanations

Why it’s dangerous:

- Sounds convincing
- Breaks trust
- High risk in finance, health, or legal use cases

How to reduce hallucinations:

- Add guardrails
- Limit scope
- Avoid asking for facts unless necessary

Rule of thumb:

> Hallucinations are when AI sounds confident but is wrong.

Summary Framework

Factor	Meaning
Cost	How expensive the agent is to run
Frequency	How often it runs
Output Quality	How usable the output is
Hallucinations	How often it makes things up

Key Insight

> Building AI agents is about **trade-offs**, not perfection.

Or the shorter version:

> We’re not trying to make AI smarter — we’re trying to make it predictable.