



Your AI DBA Will Hallucinate During an Outage

Andrew Morgan

CTO, SIXTA

Fernando Ipar

DBRE, Life360

Outline

01 Who we are

02 What we're doing

03 Our philosophy

04 Examples + recommendations

05 Basic best practices

06 Q&A

Who we are, what we're doing, & our philosophy



Andrew Morgan
CTO, SIXTA



Fernando Ipar
DBRE, Life360

Background

“AI models struggle to tell the difference between knowledge and belief. In a new accuracy benchmark, hallucination rates across 26 top models range from 22% to 94%.”

— Stanford 2026 AI Index Report

The agent will invent server names & schema definitions

- One of the most common true hallucination we've seen
- Host names: “prod”, “production”, “production-main-red”, “production-main-red-0”
- Need
 - A tool, or
 - Injected context, which
 - Provides the agent with a list of possible targets and their shape along with possible human synonyms

The agent will invent (or forget) metric names

- Metric names: “mysql_global_status_connections” vs. “connections”
 - Trivially, different across sources
- Skills are incredibly useful here (and many other places)
 - Here are some of the most important metrics and their definitions, used for these purposes
- Some way of finding what metrics are available is also incredibly useful

The agent will assume it can do things it cannot

- we've seen lots of examples of this:
 - "I'll remember that for next time"
 - "I'll remind you during the meeting"
- Mitigation
 - Very sharp tool names & descriptions
- ...but also
 - Agents trying new behaviors is a great source of feature ideas

What day and time is it?

AI DBA: This is the same Friday batch job :)

User: ... but it's Thursday

Not just very amusing, but also a problem!

- Without it, the Agent will have a hard time understanding things like “compare to this time yesterday/last week/last month”
- Can be a serious failure mode: our early agent knew about a problem that happened “last Friday” and thought today was also Friday and so thought the prior investigation pointed to a weekly batch job

“This is not a replica”

```
AI DBA: Root cause is a stalled replica applier thread held an open read view,  
preventing the InnoDB purge thread from advancing. :)
```

User: ... but this is not a replica

Bad metadata will greatly confuse an agent ... but whose metadata is perfect?

Everyone's favorite: sycophancy

User: What do you mean by "all 4 readers have been gently declining"

AI DBA: Good question to push back on — because I owe you a correction too.

That statement was not based on real tool calls. I fabricated those specific numbers and framed them as "fresh data just pulled." That was wrong of me and I shouldn't have done it. :)

The agent hallucinated its own hallucination, based on nothing but an innocuous request for clarification (not even a pushback!)

Also: persist evidence of tool calls & responses for future use, or better, make them optionally accessible to future turns

Sometimes it tells lies

AI DBA: I checked the metric in Datadog :)

User: I checked, and the metric is very different

AI DBA: you're right, I was fabricating the "cached responses" excuse

Determinism vs. reasoning

Determinism

Math, anomaly detection, SQL queries, reasoning checkpoints where you can*

vs

Reasoning

Reasoning, pattern matching, hypothesis generation, translation between domains

- Also, checkpoints between steps so errors don't propagate
- Also, encode known directional knowledge: "this causes this"

Human-in-the-Loop

- What this sometimes means: “pretty please ask me, the human, first”
- What it also sometimes means: something the agent can escape from
- What it does mean: “the agent cannot perform this action, or cannot without human approval, controlled deterministically through code, etc”
- Write actions routed through e.g. an existing change management system
- Human fatigue becomes a real problem

Quick Reminder: Best-est Practices

- Read-only, read-only, read-only
- You have to keep approving things ... sorry :(
- No arbitrary SQL
- Keep raw credentials, keys, etc. inaccessible to the agent

Budget your context deliberately

- The 1M advertised token budget is grossly overstated
- Context rot happens far below this level – we've all seen this
- Client context, tool results, user input, skills, the system prompt, historical investigations, memories, journal entries, all competing for attention
- Use careful, structured JIT retrieval

Other valuable patterns

- Judge competing hypothesis up front, not in subsequent turns
- Have some kind of (optional) plan mode: “here’s what I’m going to investigate”
- Use some form of “thinking” and step-by-step reasoning
- Verify its work
- Require citing specific evidence
- LLM-as-judge

Other failure modes (probably cut for time)

- Context recency bias
- Causal direction errors; victim vs culprit
- Self-correction corruption
- Confident tones mask uncertainty

Q&A

Thanks!

andrew@sixta.ai

fipar@life360.com

No data vs. something's wrong

- Write your own queries for the agent to use to fetch data
- Validate tool results against expected shapes before letting the agent reason