Better Prompts Can't Fix Bad Data.

Three weeks tuning prompts got marginal gains. Three months cleaning data cut hallucinations 60%.

SWIPE →

Data Habits That Actually Move the Needle

01 Treat data quality as an explicit objective, not a slogan

02 Design ingestion as a quality filter, not just a pipe

03 Curate a canonical knowledge base

04 Chunk by meaning, not arbitrary token counts

<70%

Accuracy When Models Read Noisy Data

Hallucination becomes highly prevalent on noisy, poorly curated datasets. This is a corpus problem, not an architecture problem.

Recent LLM hallucination surveys

60%

Drop in Our Hallucination Rate

After three months spent cleaning data instead of tuning prompts — without changing a single line of prompt code.

Author's own RAG deployment

89%

Answer accuracy after the data cleanup — up from 65% before, with zero changes to prompt code.

Author's own RAG deployment

"

A 70-point reasoning engine reading a 40-point corpus will not average out; it will anchor to the weakest link.

Terence Kok
Enterprise AI Strategist

Prompt Tuning vs. Data Cleaning

PROMPT TUNING	DATA CLEANING
Marginal improvements at best	Canonical knowledge base
Tweaking temperature, testing models	Ingestion as a quality filter
Polishing a broken mirror	60% fewer hallucinations

Garbage Retrieved, Garbage Reasoned.

RAG doesn't magically fix a broken knowledge base. Even with high retrieval recall, systems still fail when the retrieved documents themselves are low quality or missing key context.

The Full Data Discipline

Filter on ingest → Canonical sources → Semantic chunking →
Evaluate together → Close the loop

The full breakdown and data quality checklist:
terencekok.com/blog/stop-tuning-prompts-start-cleaning-data

The Best LLM Upgrade Wasn't a Model.

It was three months spent cleaning data. Not one line of prompt code changed.

Terence Kok · Enterprise AI Strategist

SAVE · COMMENT · FOLLOW