Complex AI != Better Solution

Learn when NOT to Use These Popular AI Techniques

PART 1

Bhavishya Pandit

The “Too Much Tech” Trap

Source

We’re living in an era where tech moves faster than you can say “fine-tune a LLaMA for my use case”.

Every week, there’s a new hype:

→ Naive RAG is dead; a new technique will fix everything!

→ Agents will do your job!

→ Fine-tune it! No, prompt it! New tools!

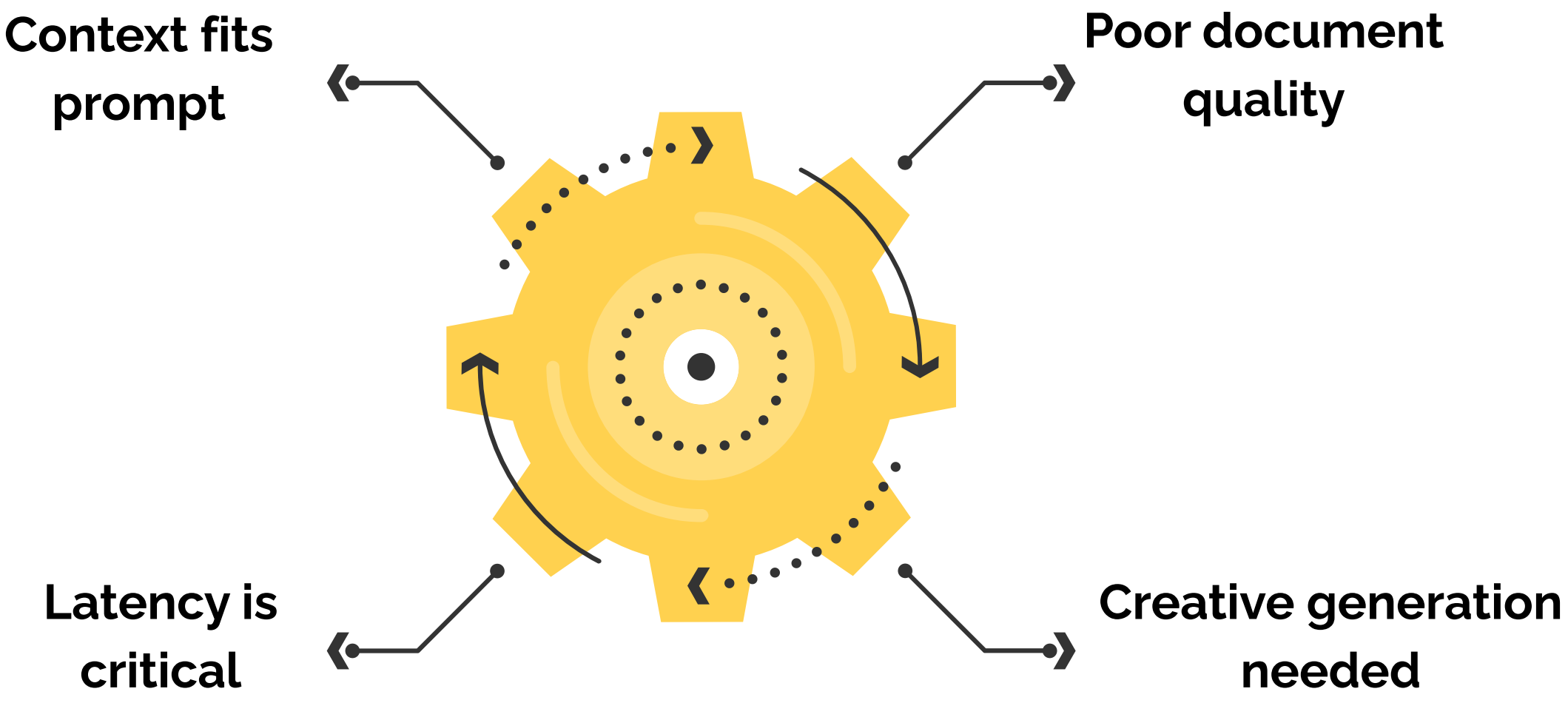
But here’s the truth:

More tech ≠ better solution.

Sometimes it is important to realise - all that glitters is not gold.

Today you will learn - when not to use certain AI techniques.

1. RAG



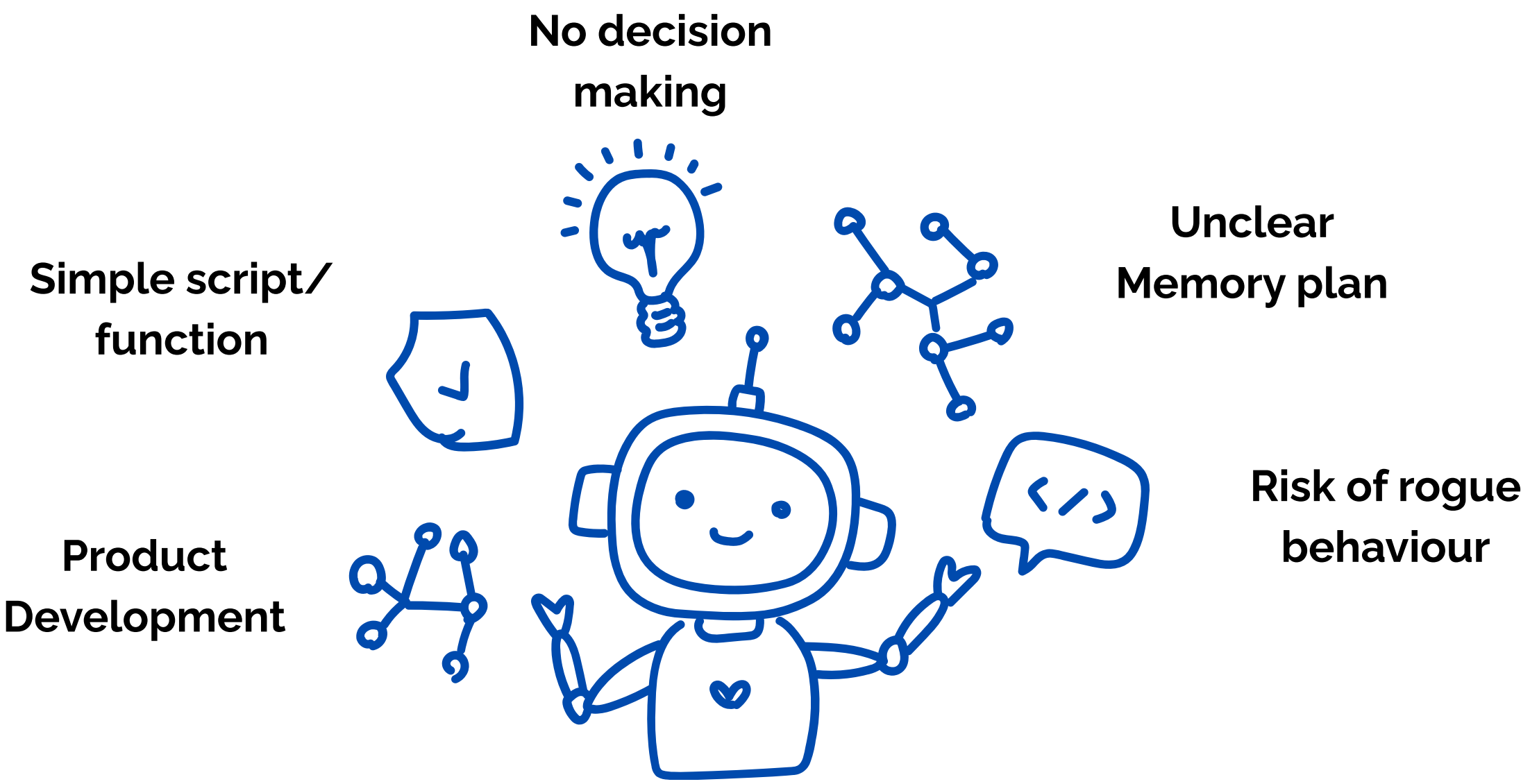
Don’t use RAG if:

- Your context already fits in the prompt window.
- You don’t have high-quality retrievable docs.
- You just need a creative or opinionated generation.
- Latency is a dealbreaker, RAG adds hops.
- You can get by with simple prompt engineering

Why?

RAG adds extra retrieval steps, needs reliable docs and infra, and slows responses if your context already fits or a smart prompt does the trick, then keep it simple.

2. AI Agents



Don’t use AI Agents if:

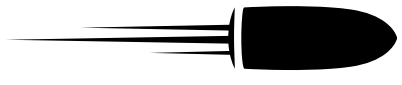
- A simple script or function does the job.
- The task has zero decision-making or chaining.
- You don’t have a clear memory/retrieval plan.
- You’re building a product, not a research project.
- You don’t want things to go rogue (and expensive)

Why?

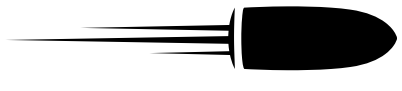
Agents can overcomplicate simple tasks, can lose context without memory plans, undermine production stability, and run unpredictably, driving up costs.

3. LLM Fine-Tuning

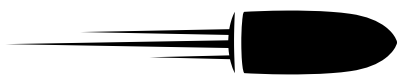
Fine-tuning sounds fancy. Custom model? Branded intelligence? Yes, please.
But hold up...



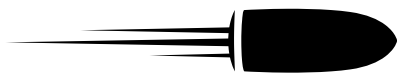
Low Data quality



High Infrastructure cost



Task not stable



Prompting solves it

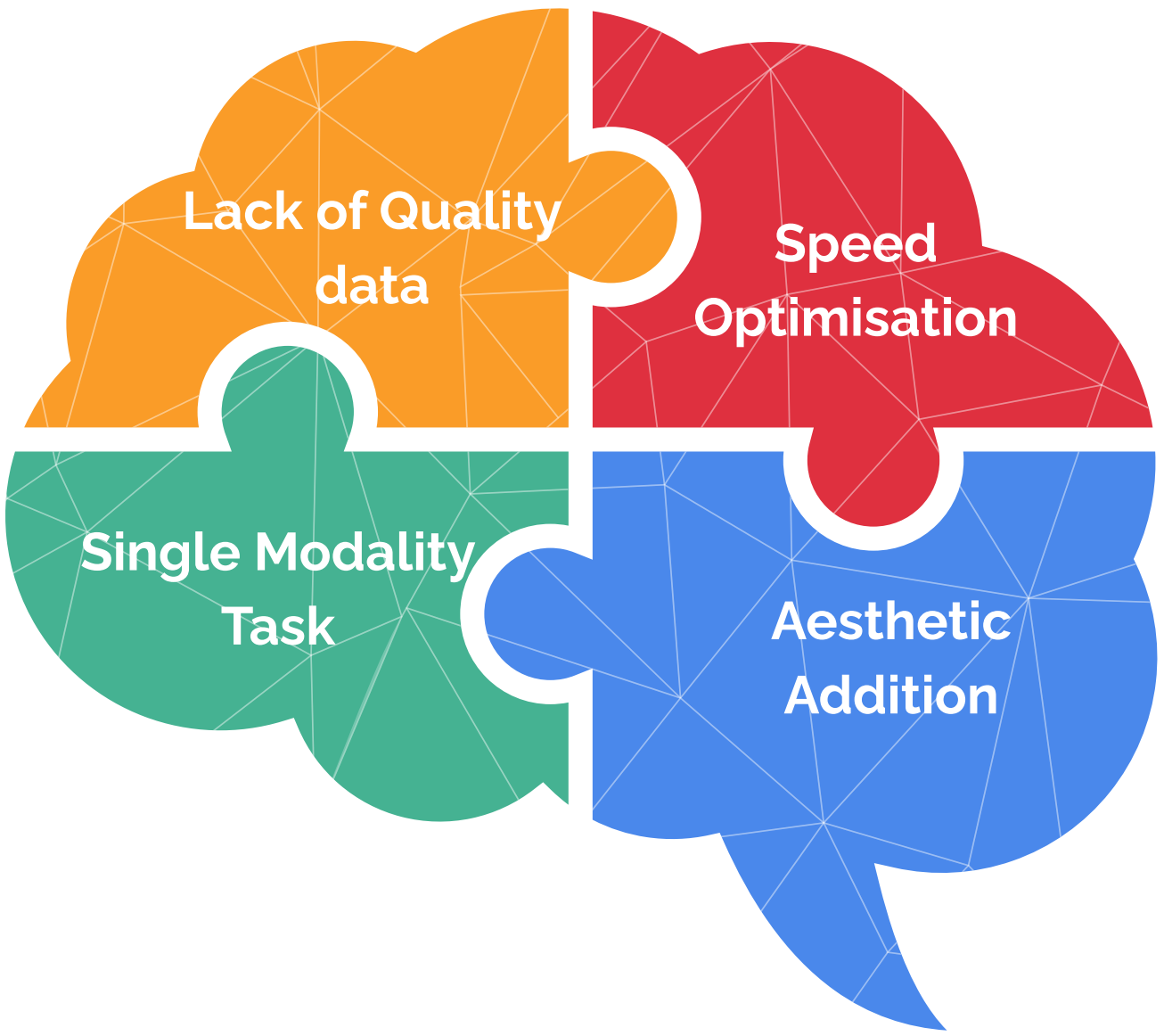
Don’t use fine-tuning if:

- Prompting or few-shot gets the job done.
- You don’t have high-quality, task-specific data.
- Your task changes often static fine-tunes age fast.
- You can’t afford the infra to serve your own model.
- Your use case isn’t complex enough to justify the cost

Why?

Fine-tuning is a big upfront lift: you need lots of good data, extra compute, and time, and it gets stale when your needs change. Often, a smart prompt or a small adapter tweak does just as well.

4. Multimodal Pipelines



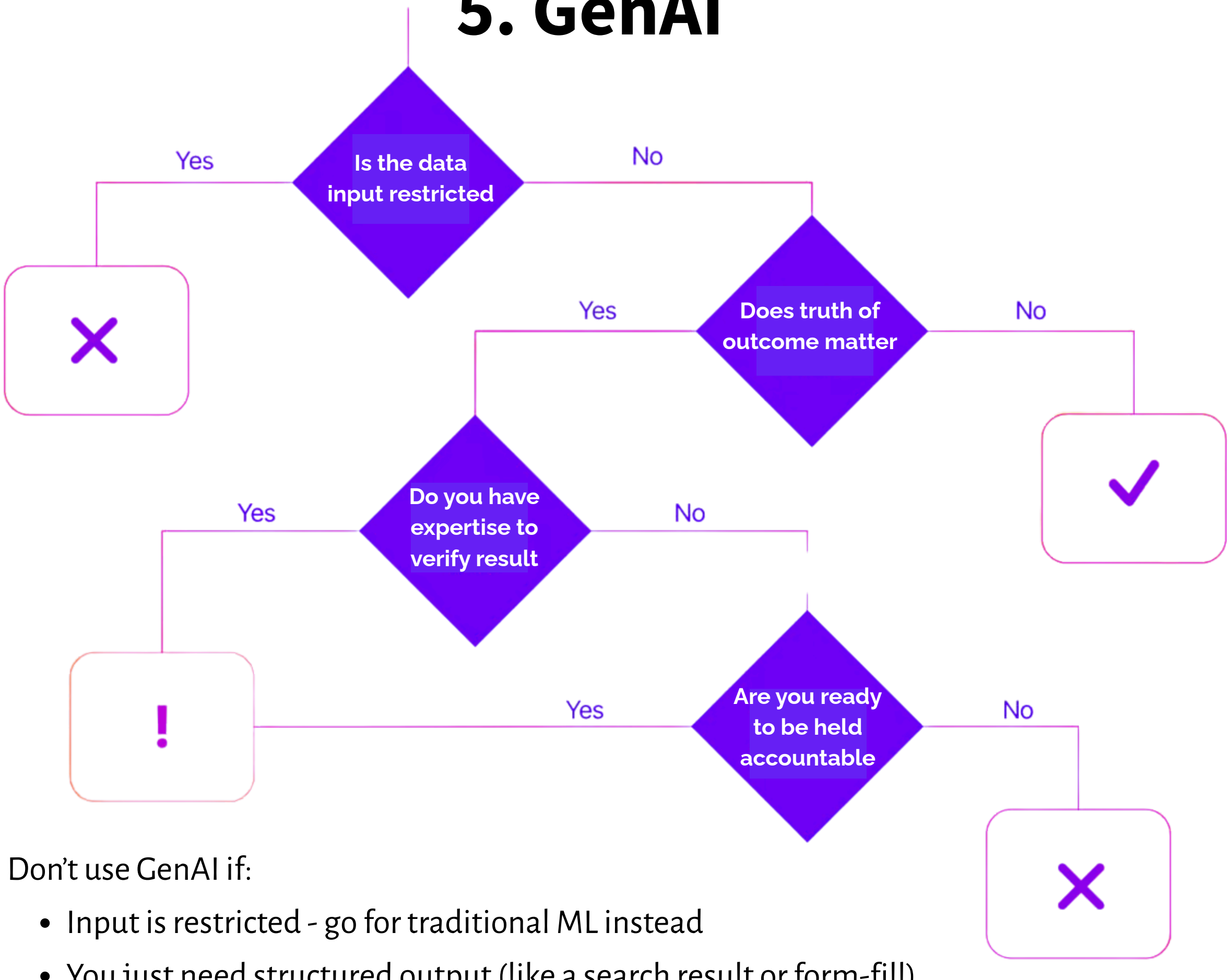
Don’t use Multimodal Pipelines if:

- Your task only needs one modality (like text alone)
- You don’t have the quality multimodal data needed for solid composition
- You’re optimizing for speed these pipelines aren’t quick
- You’re adding images or other modalities just because “it looks cool”
- You’re expecting the system to think across multiple modes like a human (spoiler: it doesn’t)

Why?

Because combining modalities in a pipeline isn’t always the best choice. Sometimes, simpler tasks don’t need the complexity of coordinating all the data especially if you can get your answer faster and cheaper with a single modality.

5. GenAI



Don’t use GenAI if:

- Input is restricted - go for traditional ML instead
- You just need structured output (like a search result or form-fill)
- You already have a fixed, human-written template that works
- You need consistent, fact-checked outputs
- You’re in a domain where hallucinations = disasters (finance, healthcare, law)
- You're just trying to “add AI” for the hype

Why?

Because GenAI is great for creativity, terrible for consistency. Use it when flexibility matters. Avoid it when correctness, control, and clarity matters the most.

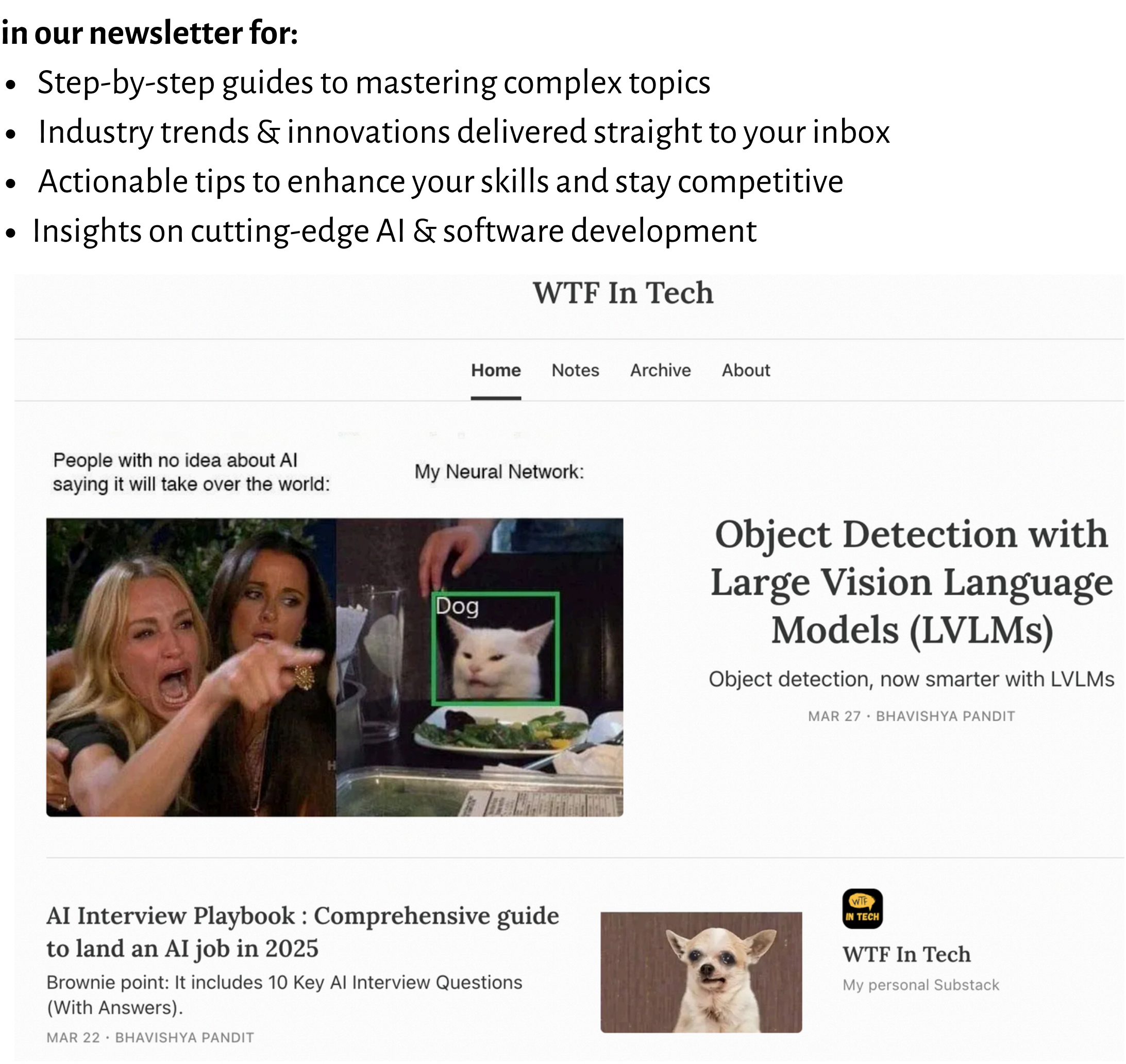
Stay Ahead with Our Tech Newsletter! 🚀

👉 **Subscribe now and never miss an update!**

🔗 **https://bhavishyapandit9.substack.com/**

Join our newsletter for:

- Step-by-step guides to mastering complex topics
- Industry trends & innovations delivered straight to your inbox
- Actionable tips to enhance your skills and stay competitive
- Insights on cutting-edge AI & software development



💡 Whether you're a developer, researcher, or tech enthusiast, this newsletter is your shortcut to staying informed and ahead of the curve.

Bhavishya Pandit

Follow to stay updated on Generative AI

SAVE

LIKE

REPOST