

Large Language Models Are an Evolutionary Dead End.

The evidence spans training data, computation, operations, and economics — not just one weak point.

SWIPE →

Why LLMs Hit a Wall

01 The scaling problem — data and compute

02 Why enterprise deployments are failing

03 The architecture that comes next

04 Economic and governance pressure

95%

Of Organisations See Zero Measurable ROI

Despite \$30–40 billion in investment, generic LLMs lack memory, contextual adaptation, and continuous improvement — the core issue is architectural, not implementation.

Source: enterprise generative AI deployment data

33–79%

Hallucination Rate in OpenAI's Newest Reasoning Models

o3 and o4-mini hallucinate more than double the rate of older o1 models. This isn't a training problem — LLMs optimise for statistical likelihood, not verified truth.

Source: OpenAI model evaluations

5%

Of custom enterprise LLM solutions ever reach production. 42% of companies abandoned most of their AI initiatives in 2025.

Source: enterprise AI implementation surveys



LLMs are trained to produce the most statistically likely answer, not to assess their own confidence.

Terence Kok

Enterprise AI Strategist

General-Purpose vs. Specialised

GENERAL-PURPOSE LLMS

Quadratic compute cost at scale

Black-box, unexplainable outputs

One model asked to do everything

SPECIALISED ARCHITECTURES

80-90% of capability, on-device

Neuro-symbolic: interpretable by design

Multi-agent, task-specific components

Retrieval Doesn't Buy **Understanding**.

RAG is often positioned as extending LLM capabilities, but it introduces its own failure modes: crude chunking, costly re-indexing on every data change, and — most critically — it doesn't fix the underlying problem. The model still generates responses probabilistically, without verifying factual accuracy, even when the right context was retrieved.

Where the Field Is Actually Heading

Domain-specific models · Neuro-symbolic AI · World models ·
Multi-agent orchestration

The full technical case, architecture by architecture:
terencekok.com/blog/why-large-language-models-future

The Future Is Specialised, Modular, and Interpretable.

Organisations still scaling general-purpose LLMs will find themselves on the wrong side of the technological divide.



Terence Kok · Enterprise AI Strategist

SAVE · COMMENT · FOLLOW