Paper 0 — Foundations of Applied Intelligence

The Four Drivers of the Intelligence Transition

Every major shift in computing begins quietly. The transistor did. The microprocessor did. The Internet did. The cloud did. What appears sudden is usually the convergence of ideas that have matured quietly for decades. Generative AI is one of those moments.

It did not emerge because the world decided it wanted “AI”. It emerged because four technical drivers matured together:

- neural networks that learn behaviour
- geometric representations of meaning - vectorisation
- the attention architecture
- industrial-scale compute.

Once aligned, they changed what machines can do.

Learning Replaces Instruction – neural networks

For most of computing history, software has been deterministic. Engineers write rules. Systems execute them. If behaviour is wrong, you edit the code.

Neural networks follow a different path. Instead of encoding logic directly, we define a flexible model and allow it to learn from data. Internally, the model contains weights — large arrays of numerical parameters. During training, those numbers are adjusted until prediction error falls. The resulting configuration determines behaviour.

Two important mathematical techniques - backpropagation and gradient descent make this adjustment possible, turning abstract models into trainable systems. At its core, modern AI is large-scale optimisation over high-dimensional parameter space.

This marks a structural shift from deterministic logic to statistical inference as the foundation of software behaviour. We are no longer specifying every rule. We build systems that infer structure from examples. Behaviour is not written line by line. It is discovered.

The revival of neural networks in the 2000s — under deep learning — was enabled by improved optimisation, large digital datasets and parallel hardware. Vision crossed commercial thresholds first. Speech followed. Language was next.

Software became adaptive.

Meaning Becomes Geometry - vectorisation

The second shift was representational.

Language and other data types can be mapped into high-dimensional vector spaces. In those spaces, similarity becomes proximity. Words used in similar contexts cluster together. Sentences, documents, images and code can be embedded into the same geometric framework.

These representations are learned rather than assigned, emerging from optimisation processes that compress patterns in large datasets into geometric structure. Self-supervised learning removed the need for manual labelling. Models learn by predicting structure within the data itself. That breakthrough made large-scale language training feasible. This replaces symbolic manipulation with structural navigation. Systems no longer depend on exact wording. They operate over relationships captured in space.

Computing moved from matching strings to modelling meaning. Meaning became navigable.

Attention Changes the Architecture - transformers

In 2017, the paper *"Attention Is All You Need"* from Google Brain introduced the Transformer architecture. It was developed to address translation but unlocked something larger.

Earlier models processed sequences step by step, compressing context as they went. Attention allowed every element in a sequence to reference every other element directly. The system learned which relationships mattered when generating output.

The Transformer architecture aligned naturally with parallel hardware, allowing models to scale without sequential bottlenecks. Attention turned neural networks into general context processors.

Large language models became possible. These principles now power systems that generate language, images, audio and increasingly video.

Scale Makes It Real – datacentres and the internet

The final driver is scale. These ideas existed for years. They transformed the industry only when deployed at industrial magnitude.

Training modern models requires vast GPU clusters, high-bandwidth networking, specialised data centres and significant capital investment. Training runs depend on massively parallel matrix computation refined over decades and hardware designed for it. Companies such as NVIDIA, Microsoft, Google and Amazon built the datacentre infrastructure to support it.

The modern Internet provides the other essential ingredient. It creates a machine-readable corpus of human knowledge at planetary scale and a global distribution layer for deploying models as services. Without this digitised substrate and networked delivery, large-scale training and real-time inference would not have been possible.

Models have expanded from millions to billions and then hundreds of billions of parameters. As they grew, performance improved predictably. Capability scaled not linearly but qualitatively. New behaviours emerged from quantitative growth. Capabilities that once seemed aspirational became operational.

Training frontier models now requires billions of dollars in infrastructure, concentrating capability in organisations able to finance and operate at that scale. Intelligence at this scale is no longer purely a software challenge. It is an energy and infrastructure challenge. At scale, cognitive capability depends on infrastructure.

Compute is now a capital asset class.

Abstraction - and its consequences

This transition also marks an abstraction shift.

Computing advances by hiding complexity. Hardware became operating systems. Operating systems became cloud platforms. Each layer increased leverage by reducing lower-level detail – this is abstraction.

AI abstracts behaviour itself. Instead of writing logic, we define objectives. Instead of defining relationships, we embed them. Instead of specifying relevance, the system learns it.

But abstractions lose sight of the underlying processes. They can cause failures – and this led to the term leaky abstractions.

When underlying constraints reassert themselves, failure surfaces at the higher layer. A language model appears to reason — until it hallucinates. A retrieval system appears robust — until context drifts. An API feels lightweight — until infrastructure limits intrude.

The more capability we abstract, the more we depend on layers we do not directly understand or control. This is not a flaw. It is structural.

The Intelligence Transition

When learning systems, geometric representation, attention and scale converge, computing changes character. Software moves from explicit instruction to probabilistic modelling. Systems shift from tools that execute commands to platforms that infer intent.

Generative AI is not a feature upgrade. It is a platform shift. This shift was not triggered by marketing. It emerged from mathematics, architecture and infrastructure reaching sufficient scale. Once those drivers aligned, the outcome was not optional. It was inevitable.

We have crossed a threshold where machines no longer simply execute code. They model structure, generate language and assist in reasoning. That changes how organisations operate. It changes how infrastructure is built. It changes how value is created. Once systems learn from data at scale, reverting to purely deterministic software is economically irrational.

The Intelligence Transition is underway. The consequences extend beyond technology. They reshape markets, capital flows and organisational design.

Alan Moore, 3 March 2026