The Kardashev Scale Holds a Hidden Lesson for AI

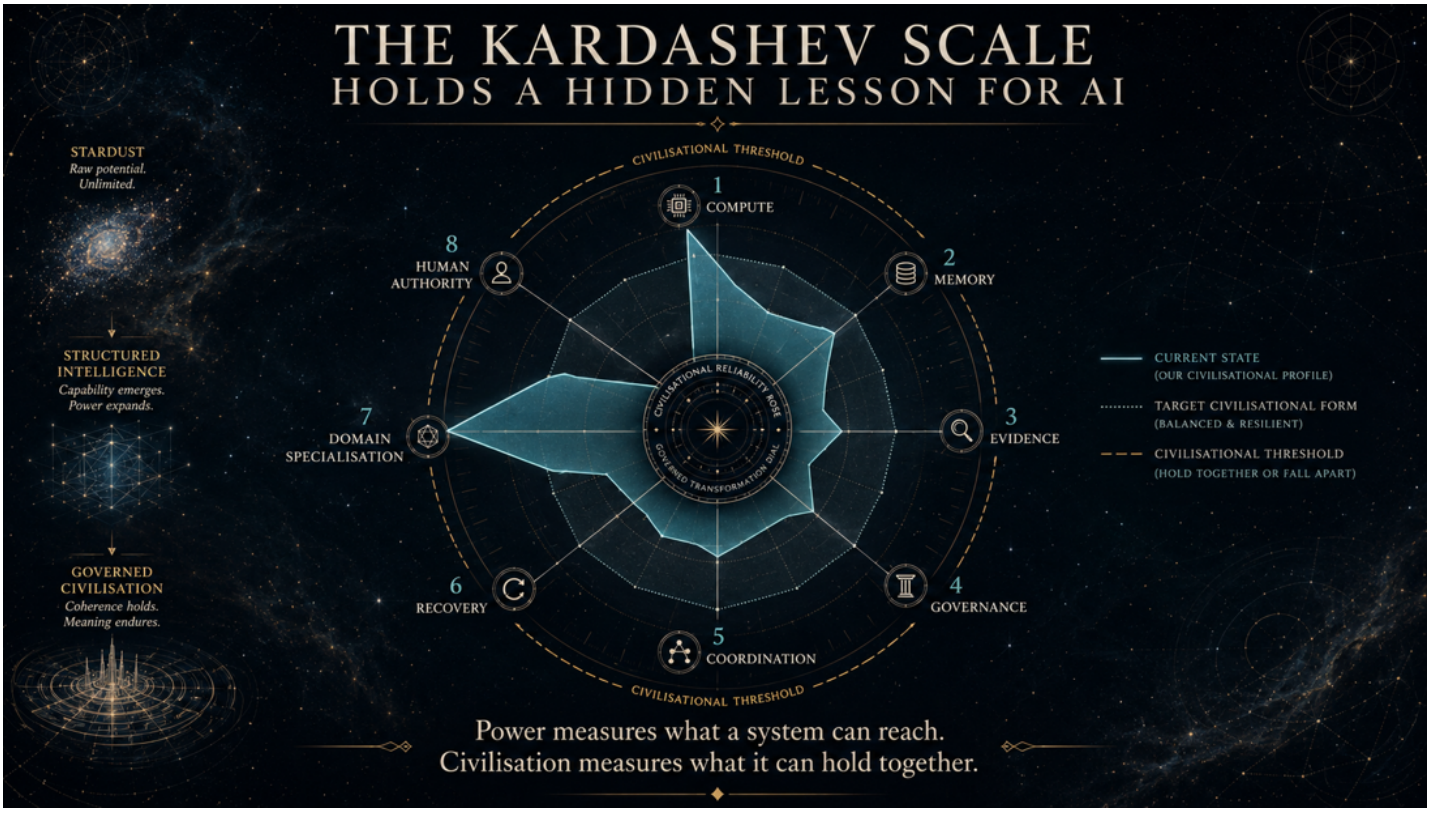
From: **Joseph.Maxwell**
Joseph.Maxwell02@proton.me

To: **Joseph.Maxwell02@proton.me**
Joseph.Maxwell02@proton.me

On: **Thursday, 18 June 2026 at 19:59:57**

Subtitle:

AI already contains the stardust of a universal intelligence. The next challenge is turning raw capability into a governed civilisation.



Artificial intelligence already contains something extraordinary.

Inside a modern language model sit compressed traces of mathematics, engineering, biology, law, programming, literature, design, history and almost every other field humans have managed to express through language.

It is tempting to call that a universal intelligence.

But that would be premature.

A cloud of stardust may contain the material from which stars, planets, ecosystems and civilisations eventually form. That does not mean the cloud is already any of those things.

The difference is structure.

The same underlying matter can remain dispersed across space, collapse into a star, assemble into a planet, enter a living organism or become part of a civilisation capable of observing the universe that produced it.

The material may be shared.

The organisation is not.

Artificial intelligence is in a similar position.

The raw material is present.

The architecture is not.

We have created systems with access to an enormous field of possible capability, yet we still tend to treat them as single, undifferentiated minds.

We ask the same model to write code, interpret science, organise evidence, analyse policy, advise on decisions, plan experiments and explain its own conclusions.

It attempts all of those tasks through one conversational surface.

That is not a civilisation.

It is stardust being asked to behave like a city.

The next stage of artificial intelligence may therefore depend less on making models larger and more on giving their intelligence structure.

That means specialisation.

Boundaries.

Memory.

Evidence.

Coordination.

Recovery.

Authority.

And a governing system capable of holding all those parts together without allowing any one capability to become the whole machine.

Strangely enough, one of the clearest fictional images for this architecture comes from Ben 10.

Yes.

The Omnitrix.

The Omnitrix problem

The Omnitrix does not simply give its wearer unlimited power.

It contains a library of specialised forms.

Each form has its own capabilities, limits, strengths, weaknesses and operating conditions.

One form may be useful for speed.

Another for strength.

Another for intelligence.

Another for surviving a hostile environment.

The wearer does not become all of them at once.

A form is selected.

The transformation is bounded.

The identity of the operator remains.

There are time limits.

There are emergency reversions.

There are restrictions on access.

The system can refuse.

The value of the Omnitrix is not only in the powers it contains.

The value is in the system that governs their use.

That is much closer to the architecture intelligent software now needs.

A model working in materials science should not operate under the same memory, tools and evidence rules as one working in cosmology.

A coding system should not automatically inherit authority to promote scientific claims.

A documentation engine should not be able to alter experimental thresholds simply because it can describe them.

A project organiser should not become a model validator.

The problem is not that these capabilities cannot coexist.

The problem is that they must not collapse into one another.

A mature intelligent system should be able to identify the domain, select the valid form, load the correct memory, activate the permitted tools, apply the appropriate constraints and then return to the governed core.

The Governed Transformation Path

1. Intake
   A request enters the system.
2. Classification
   The domain, authority and risk are identified.
3. Transformation
   The correct domain capsule is loaded.
4. Isolation
   Only the relevant tools, memory and evidence rules become available.
5. Execution
   The task runs inside declared boundaries.
6. Audit
   Evidence, limitations and failures are recorded.
7. Reversion
   The system returns to its governed core.

The Omnitrix gives us the local picture.

It helps us imagine how one system might hold many specialised forms.

To understand what happens as those forms scale into entire networks of intelligence, we need a second metaphor.

That is where the Kardashev scale becomes useful.

Kardashev's beautiful simplification

The Kardashev scale is one of the most elegant thought experiments in science.

Rather than trying to list every technology an advanced civilisation might develop, it asks a simpler question:

How much energy can that civilisation harness?

A Type I civilisation operates at roughly planetary scale.

A Type II civilisation operates at stellar scale.

A Type III civilisation operates at galactic scale.

The scale does not attempt to predict the exact machines, institutions or cultures such

civilisations would possess.

It measures the resource envelope they can coordinate.

That is what makes it powerful.

It reduces an almost unimaginable variety of possible civilisations to a simple scaling relationship:

resource access
→ coordination capacity
→ civilisational scale

The point is not that energy alone defines civilisation.

The point is that the scale asks how much power a system can bring under organised control.

That logic can be translated.

Not literally.

Structurally.

The Analogia gate

A weak analogy maps objects by appearance.

A strong analogy maps mechanisms.

So we should not say:

A company is a planet.

A data centre is a star.

The cloud is a galaxy.

Those are surface substitutions.

They may sound impressive, but they explain very little.

The real mechanism inside the Kardashev scale is this:

A civilisation advances when it can access a larger resource envelope and coordinate that resource without collapsing under the burden of its own scale.

That mechanism can be transferred into artificial intelligence.

The mapping becomes:

Energy capacity

→ computational and cognitive capacity

Energy distribution
→ routing of tasks, models, memory and evidence

Civilisational infrastructure
→ shared runtime, contracts, tools and governance

Planetary or stellar coordination
→ organisational or distributed coordination

Energy misuse
→ capability outrunning authority

This is an example of what I call Analogia.

The goal is not to borrow the imagery of another field.

The goal is to isolate the invariant relationship beneath it and test whether that relationship remains useful when transferred elsewhere.

The invariant here is:

resource scale
→ coordination burden
→ governance requirement
→ civilisational capability

The AI translation is:

intelligence scale
→ coordination burden
→ governance requirement
→ operational civilisation tier

Analogia decides whether the transfer is legitimate.

Systematica Mathematica is the broader method of representing the relation that survives the transfer.

The names matter less than the discipline:

find the mechanism, preserve the structure, declare what does not carry across.

AI already has the raw material

A foundation model is not simply a tool.

It is closer to an undifferentiated field of possible tools.

Within the same model may sit the capacity to:

write software,

summarise research,

reason about engineering,

analyse documents,

generate designs,

translate languages,

plan experiments,

organise projects,

or explain mathematical ideas.

That breadth is astonishing.

But breadth is not the same as organisation.

A civilisation is not defined by one entity attempting every possible role.

It is defined by differentiated systems that can work together.

A hospital is not one person pretending to be every specialist.

A city is not one building containing every function.

A scientific field is not one researcher attempting every experiment.

Civilisation appears when specialisation, memory, authority and coordination become stable enough to support collective action.

Artificial intelligence currently has much of the material.

It has models.

It has compute.

It has tools.

It has data.

It has memory systems.

It has autonomous workflows.

It has human operators.

But these elements are often connected loosely.

A chat history becomes memory.

A tool call becomes evidence.

A fluent explanation becomes authority.

A temporary interpretation becomes doctrine.

A model prediction becomes a decision.

The problem is not a lack of intelligence.

The problem is that the intelligence remains structurally primitive.

From raw intelligence to usable civilisation

We can begin with a simple model.

Let the raw capacity of an intelligent system be represented as:

$$R_{raw} = w_C \cdot C + w_M \cdot M + w_D \cdot D + w_T \cdot T + w_H \cdot H$$

Where:

C = compute
M = memory
D = accessible data
T = tools and specialised modules
H = participating human intelligence

The weights indicate that these resources do not naturally share the same units. They must be normalised for the system being examined.

This is not yet an empirically calibrated law.

It is a first operational model for separating available capacity from governable capacity.

Raw resources still do not tell us how much intelligence the system can use safely and coherently.

A system may have enormous compute but poor coordination.

It may have vast memory but weak provenance.

It may have many tools but no clear authority boundaries.

It may have several models but no reliable way to resolve disagreement.

So we need a second expression:

$$R_{civil} = R_{raw} \times \eta_g \times \eta_c \times \eta_e \times \eta_m \times \eta_r$$

Where:

ηg = governance efficiency
ηc = coordination efficiency
ηe = evidence integrity
ηm = memory coherence
ηr = recovery capacity

Each efficiency term lies between zero and one.

The first equation estimates what resources are available.

The second estimates how much of that resource can be used coherently.

This creates a basic but important result:

More raw intelligence does not necessarily create more usable intelligence.

Imagine two systems.

System A has moderate raw capacity, but strong governance, evidence, memory and recovery.

System B has far greater raw capacity, but weak controls and poor recovery.

For illustration:

System A:

$R_{raw} = 100$

ηg = 0.9
ηc = 0.8
ηe = 0.9
ηm = 0.8
ηr = 0.7

$R_{civil} \approx 36$

System B:

$R_{raw} = 250$

ηg = 0.3
ηc = 0.4
ηe = 0.5
ηm = 0.5
ηr = 0.2

$R_{civil} \approx 1.5$

The second system possesses much more raw capability.

It has far less coherent operational intelligence.

That is the difference between having more power and being able to organise it.

The weakest-link law

There is another problem with averaging.

A system with excellent compute, memory and tools may still fail because one critical layer is weak.

So we can define a civilisational bottleneck:

$B = \min(\eta g, \eta c, \eta e, \eta m, \eta r)$

The effective tier of the system is limited by the weakest indispensable dimension.

This gives us a useful principle:

A system with galactic compute and local governance remains a local-grade civilisation carrying galactic risk.

That line matters.

It explains why raw capability should never be confused with maturity.

A system does not become more civilised simply because it can reach further.

It becomes more civilised when it can hold more power together without losing coherence, evidence or control.

A civilisation scale for intelligent systems

If we borrow the scaling logic of Kardashev, we can imagine a civilisational scale for intelligent software.

Not as a literal scientific classification.

As an operational model.

Type 0: Powerful fragments

At Type 0, capability exists, but it is not coherently organised.

There may be isolated scripts, separate chat histories, messy project folders, overlapping models, temporary memory, unclear authority and no stable evidence state.

This describes much of present-day AI use.

The system may be powerful.

It is not yet civilisational.

Type I: Local civilisation

A Type I intelligent civilisation operates as one coherent environment.

It may include one human operator, one or more models, several specialised modules, persistent memory, a shared evidence standard, clear authority boundaries and local recovery.

The defining feature is not scale.

It is unity.

The system can act across several functions without losing track of what ran, what was observed, what was inferred and who had authority to decide.

Type II: Federated civilisation

At Type II, multiple teams, projects or organisations operate under compatible governance.

Memory and compute may be distributed.

Local systems retain sovereignty.

Evidence moves through governed contracts.

The challenge is federation without flattening.

Type III: Distributed civilisation

At Type III, the intelligent estate becomes too large for any single node to inspect directly.

Different infrastructure, institutions, models and time horizons must still cooperate.

Failures must remain local.

Evidence must remain traceable.

Authority must be distributed.

Type IV: Universal civilisation

A Type IV system coordinates across many domains and technical environments.

Materials science, biology, economics, engineering, documentation, software and design may all participate in the same governed environment.

But they must not be reduced to one mechanism.

Universal does not mean uniform.

It means interoperable without erasure.

Type V: Multiversal civilisation

At Type V, the system can maintain multiple internally coherent but potentially incompatible universes.

Each may contain its own assumptions, ontology, memory, tools, models, evidence rules and authority structure.

These universes can exchange bounded packets.

But no universe automatically inherits another's truth.

A concept moving from cosmology into materials science does not arrive as validated knowledge.

It arrives as a candidate mapping.

It must be scanned, quarantined, translated and tested locally.

Cross-universe transfer therefore requires a leakage audit.

The target domain must not silently inherit the source domain's assumptions, vocabulary or authority.

This is not merely multidisciplinarity.

It is the ability to preserve multiple worlds without forcing them into premature synthesis.

And to be clear, civilisation here does not mean centralising every intelligence under one mind.

It means allowing specialised systems to coordinate through shared constraints while retaining local identity, memory and authority.

Krato-Etho: governing power before it becomes action

The Kardashev analogy gives us a clean way to think about power.

Every intelligent system contains a difference between what it can do and what it should be allowed to do.

That distinction is easy to blur.

A model may have the capability to alter a file.

That does not mean it has authority.

A model may be able to generate a scientific claim.

That does not mean the evidence supports publication.

A system may be able to activate a tool.

That does not mean the action is reversible or safe.

This is where Krato-Etho enters.

Kratologia asks:

What power is reachable?

Ethologia asks:

What power is admissible?

Governance asks:

What power can be coordinated coherently?

We can express that as:

P_reachable ≥ P_admissible ≥ P_operational

Capability says what could be done.

Ethics and authority say what may be done.

Governance says what can be done coherently.

A mature system should be able to possess greater capability than it is currently permitted to use.

That is not a limitation.

It is civilisation.

The language model therefore should not sit above the whole architecture as its ruler.

It should sit inside the system as one powerful but bounded cognitive component.

The model may be the most articulate part of the machine.

That does not make it sovereign.

The role of memory

Civilisation depends on memory.

But memory is also one of the largest sources of contamination.

If a system stores every interpretation, analogy, draft and confident sentence in one undifferentiated memory, old narratives begin to influence new reasoning.

A speculative idea becomes precedent.

A temporary explanation becomes doctrine.

A mistaken summary becomes substrate.

So intelligent memory must be structured.

A mature system should distinguish between:

substrate memory,

structural memory,

interpretive memory,

narrative residue,

and temporary working cache.

Substrate memory records what entered the system.

Structural memory records relationships, states, versions and evidence.

Interpretive memory records hypotheses and explanations.

Narrative residue records presentation-specific language.

Temporary cache records transient working material.

These categories should not have equal authority.

This is another place where the stardust metaphor matters.

Civilisation does not emerge by preserving every particle in one pile.

It emerges by giving different materials different roles.

The role of the Omnitrix

The Omnitrix becomes the mechanism for switching between domain civilisations.

Turning the dial should not merely activate a different calculation.

It should load an entire governed domain capsule:

the correct engine,

the correct memory,

the correct tools,

the correct schemas,

the correct evidence rules,

the correct vocabulary,

the correct risk limits,

and the correct claim ceiling.

A materials capsule should not expose cosmological memory by default.

A coding capsule should not inherit scientific authority.

A project-intelligence capsule may organise evidence without being allowed to modify experimental results.

Each domain stands alone.

Yet all can participate in the same larger civilisation.

That is the real power of the Omnitrix analogy.

It shows how universal capability can exist without universal blending.

From metaphor to software

This is not only a philosophical exercise.

It points toward a different kind of software architecture.

One in which:

a governance layer controls the system,

a Krato-Etho layer limits reachable power,

a perceptual layer prevents weak evidence from appearing stronger than it is,

a memory layer preserves state without allowing narrative contamination,

an Omnitrix layer switches between domain capsules,

and specialised modules perform the actual scientific, technical or organisational work.

The language model becomes the operator interface.

Not the ruler.

The first practical modules might organise messy project estates, audit model reliability,

map structural regimes or help narrow experimental search spaces.

The long-term objective is not to build one model that pretends to be every specialist.

It is to build an environment where many forms of intelligence can cooperate without losing their boundaries.

That is how the product architecture mirrors the civilisation metaphor.

The software is not organised into tiers because larger customers pay more.

It is organised into tiers because larger intelligent estates require stronger coordination, memory, evidence and recovery.

The commercial structure follows the systems logic.

Not the other way around.

The threshold between power and civilisation

AI may already contain the raw materials of something civilisational.

But raw material is not architecture.

Stardust does not become a city merely because every necessary atom is present.

It must pass through structure, constraint, specialisation, memory, failure and recovery.

The same is true of intelligence.

The next generation of AI systems will not be defined only by how much they know or how many tasks they can perform.

They will be defined by how much differentiated intelligence they can hold together without allowing power to outrun evidence, memory to become contamination or capability to become authority.

The Kardashev scale asks how much energy a civilisation can command.

The question for AI is slightly different.

How much intelligence can a system coordinate without ceasing to be governed?

That is the threshold between a powerful model and an intelligent civilisation.

The material is already here.

The work now is to give it form.