# Beyond AGI: The Case for Artificial Local Intelligence

*A clear architecture for autonomous agents — without the promises nobody can keep*

Wolfgang Stegemann

Artificial General Intelligence, or AGI, means something different to everyone. The one thing most definitions share is the word general: an intelligence without a fixed domain, capable of anything a human can do.

This generality is the wrong goal for the autonomous systems. The robot that inspects an offshore wind turbine does not need general intelligence. The monitoring system that maintains a server cluster does not need to write poetry. The autonomous agent that manages a production process does not need a theory of mind.

What these systems need is a clear architecture, a defined scope of action, and a principled basis for autonomous behavior. That is what Artificial Local Intelligence, or ALI, provides.

## 1. A clear alternative to a fuzzy concept

AGI inherits the vagueness of general intelligence, which has never been satisfactorily defined. That vagueness works for investor pitches. It does not work for engineering.

ALI is a different concept entirely. An ALI is locally bounded, purpose-specific, self-preserving, and normatively constrained. It operates within a defined domain, pursues a given task, maintains the conditions for its own operation, and shuts itself down when those conditions can no longer be met. Its behavior is bounded by norms encoded in an architectural layer with hard veto power.

These are the constitutive properties of a different kind of system. An ALI that acted outside its domain would be something else, and something without the theoretical foundation that makes ALI coherent.

## 2. A defined field of action

The scope of ALI is local autonomous action in bounded domains. The inspection and maintenance of physical infrastructure requires human presence in dangerous and expensive environments. Autonomous process management has a large gap between what is possible and what is deployed. Cyber-physical monitoring could dramatically reduce response times and failure rates with autonomous agents. None of these domains requires general intelligence. All of them require reliable autonomous action within defined parameters.

ALI is designed for exactly this.

## 3. A three-layer architecture with a clear derivation

Multi-layer architectures for autonomous agents exist. BDI systems have three layers. Homeostatic agents have self-preservation. Safe AI has normative constraints. The objection writes itself.

In those architectures, the layers are design choices. In ALI, the layers are derived. They follow from a single founding principle: autonomous action in biological systems arises from operative self-preservation dynamics, and this structure can be reconstructed functionally. The causal core is the functional equivalent of what generates agency in self-maintaining systems. The Ego and Super-Ego follow from it.

This derivation changes the status of controllability. In top-down architectures, controllability is an afterthought: a constraint applied to a system optimised for performance. The tension between capability and control is structural and permanent. In ALI, controllability follows from the founding principle. A system that preserves itself through normatively constrained action cannot coherently pursue unconstrained expansion. The Super-Ego is architectonically necessary, not protective.

The architecture of ALI has three layers, each with a precise functional role.

**The causal core** is the operative self-preservation principle. It is derived from what biological self-maintaining systems do: generate gradients that constitute the physical basis of agency. It can be multi-dimensional, with an energy dimension and a structural integrity dimension that may conflict. Every action the agent takes is evaluated against it. There is no external reward. The system knows from its own state whether it is functioning.

**The Ego** is the task-executing layer. It mediates between the causal core and the environment. It can be hardcoded or learning, adapting its strategy from experience evaluated against the causal core.

**The Super-Ego** is the normative layer. It encodes the boundary of permissible action and has hard veto power. It can trigger controlled self-shutdown and can adapt its normative weights over time, generating new norms from repeated operational gaps, subject to one condition: no adaptation that destabilizes the causal core is permitted. The Super-Ego is constitutive of what the system is. Controllability follows from the architecture.

## 4. What the simulation has demonstrated

This architecture has been implemented and tested across ten simulation stages, all publicly documented. The system shows emergent self-preservation without explicit programming. It learns task execution from

internal feedback without external reward. Its normative layer strengthens rather than erodes over time. It generates new norms from operational gaps while preserving system stability. Its decisions are evaluated against a four-dimensional stability criterion integrating self-preservation, systematic consistency, normative compliance, and contextual appropriateness.

None of this required consciousness. The system has no inner life. It is an autonomous agent whose autonomy is grounded in a self-preservation principle and whose normative constraints are architecturally necessary.

## 5. Why this matters

Large language models are capable at pattern matching and generation, but they are not autonomous agents. They do not maintain themselves or act without prompting. Robotic systems are increasingly capable in physical domains, but their safety properties are typically add-ons rather than architectural foundations.

ALI acts from an internal imperative rather than an external instruction. It is controllable because controllability follows from building an autonomous agent from a self-preservation principle rather than from a capability specification.

The simulation is complete. The architecture is documented. The next step is deployment in a real domain with real sensors, real actuators, and real operational requirements.

All ten simulation stages, with papers and code:

*Stegemann, W. (2026). From the Myth of AGI to the Architecture of a Controllable ALI. https://doi.org/10.5281/zenodo.20378675*

*Stegemann, W. (2026). Four-Dimensional Decision Stability. https://doi.org/10.5281/zenodo.20407218*

*Stegemann, W. (2026). Norm Sensitisation Under Meta-Learning. https://doi.org/10.5281/zenodo.21563934*

*Stegemann, W. (2026). Assimilation and Accommodation in the ALI Super-Ego. https://doi.org/10.5281/zenodo.21564176*

*Stegemann, W. (2026). Second-Order Meta-Learning and Factorizability. https://doi.org/10.5281/zenodo.21564364*

*Simulation code (all ten stages):*
*https://github.com/drwolfgangstegemann-sudo*

Stegemann, W. (2026). Why Artificial Life and Consciousness Are Impossible in Principle. Zenodo.
https://doi.org/10.5281/zenodo.21829358

**Wolfgang Stegemann**
Independent researcher in philosophy of mind and consciousness theory
ORCID: 0009-0000-1196-1170