From Intelligence to Autopoiesis

Rethinking Artificial Intelligence through Systems Theory

Benedikt Zönnchen

11th of November 2025

MUC.DAI
Munich Center for
Digital Sciences and AI

HM

1. Why Luhmann's Systems Theory?

2. Finding an Enlightened Ground

3. Breaking the Ground

4. Artificial Systems – A Paradox?

1. Why Luhmann's Systems Theory?

2. Finding an Enlightened Ground

3. Breaking the Ground

4. Artificial Systems – A Paradox?

Systems Theory?

Why am I interested in *systems theory*?

- It makes so much sense.
- Systems thinking increasingly shapes discussions in the field of sustainability.
- It offers an abstract vocabulary that feels familiar to a computer scientist.
- Strong influence of Maturana & Varela.
- Reality made less and less sense to me $\Rightarrow$ Is my model wrong?
- Unsatisfied with the current "AI hype"
  - What *is* intelligence, superintelligence, AGI?
  - Why do we suddenly believe (again) that computers can think or even possess consciousness?
  - How do we distinguish between my autonomy and agency and that of the "tools" I use?
  - Is there a *conceptual* or *fundamental* difference between a clock and a large language model?
  - If so, what is it?

"The ability of an agent to achieve goals in a wide range of environments."
– (Russell, 2019)

Let E be the space of all computable reward summable environmental measures with respect to the reference machine $\mathcal{U}$, and let K be the Kolmogorov complexity function. The expected performance of agent π with respect to the universal distribution $2^{-K(\mu)}$ over the space of all environments E is given by,

$$\Psi(\pi) := \sum_{\mu \in E} 2^{-K(\mu)} V_\mu^\pi.$$

We call this the **universal intelligence** of agent π (Legg & Hutter, 2007).

What is Intelligence?

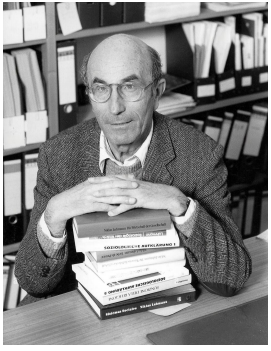
"Meine These ist, dass es keine Künstliche Intelligenz gibt. [...] Intelligenz und Denken sind die Fähigkeiten eines Lebewesens, die prinzipiell nicht in Systemen nachgebaut werden können, die nicht lebendig sind. Lebendigkeit ist eine notwendige Voraussetzung für das Vorliegen von Intelligenz."
– (Gabriel, 2024)

"Dieses Szenario, [der autopoietischen Programme, die sich selbst optimieren], setzt voraus, dass die Systeme der Künstlichen Intelligenz überhaupt intelligent sind. Sonst könnten sie das Vergleichsmaß nicht ansetzen. Sie könnten nicht sagen: 'Die sind in irgendetwas intelligenter als wir.'" – (Gabriel, 2024)

"Die Gewissheit der Unterscheidung zwischen Fiktion und Realität, die man scheinbar ganz einfach zieht, erodiert langsam. Ich verliere immer mehr von dieser westlichen Gewissheit, die mich die KI als tolles Instrument hat sehen lassen—aber sicher nicht als ein Gegenüber. Ich glaube, diese Auffassung ist falsch! Die KI ist ein echtes Gegenüber, aber natürlich kein menschliches Gegenüber." – (Gabriel, 2025)

Why Luhmann?

"Menschen können nicht kommunizieren, nicht einmal ihre Gehirne können kommunizieren, nicht einmal das Bewußtsein kann kommunizieren. Nur die Kommunikation kann kommunizieren."
– Luhmann (1927-1998)



Luhmann provides a framwork to talk about living, psychic and social system. Why not include *artificial systems*?

Why Luhmann?

1. Influeced by Hegel, Husserl, Kant, Maturana & Varela, von Foerster, Bateson, Spencer-Brown
2. Controversial with the dominant Frankfurt School (Adorno, Horkheimer, Marcuse, Habermas)
3. A sociologist who is not interested in humans (**anti-humanist**)
4. Describes but gives no normative guidance (value-neutral)
5. Often called conservative but his theory is radical
6. Hard to read (there is no starting point)
7. Excessive abstractions and obscurity, hard to falsify, maybe too grand of a theory
8. Calls himself “operational contructivist”

"It's all wrong but of high quality." – *Jürgen Habermas*

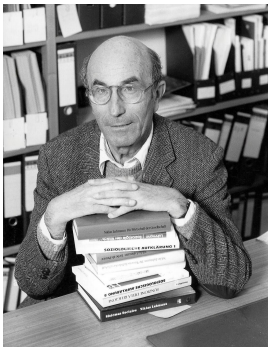
Some of Luhmann's provocations:

1. **Difference** not identity is the fundament of theory.
2. Only social systems communicate, **humans do not communicate!**
3. Once entered into communication one can no longer go back to the prardise of the soul.
4. Every system (e. g. mass media, science, sociology) observes with an unavoidable **blind spot**.
5. Descriptions seem to be made from outside but they are in fact made from within which always leads to a **paradox**.
6. Reduction of complexity brings about an increase in complexity.
7. Mass media (books, TV, social media) do not distort **reality**, they **construct** it. They conduct how society sees itself.
8. The more information, the more misinformation.

Luhmann wanted to give a complete description of society.
This includes his description ($\Rightarrow$ a paradox).

Buchtitel:

- Die Wissenschaft der Gesellschaft (Luhmann, 1992)
- Das Kunst der Gesellschaft (Luhmann, 1997)
- Die Realität der Massenmedien (Luhmann, 2004)
- Die Gesellschaft der Gesellschaft (Luhmann, 1998)
- ...



1. Why Luhmann's Systems Theory?

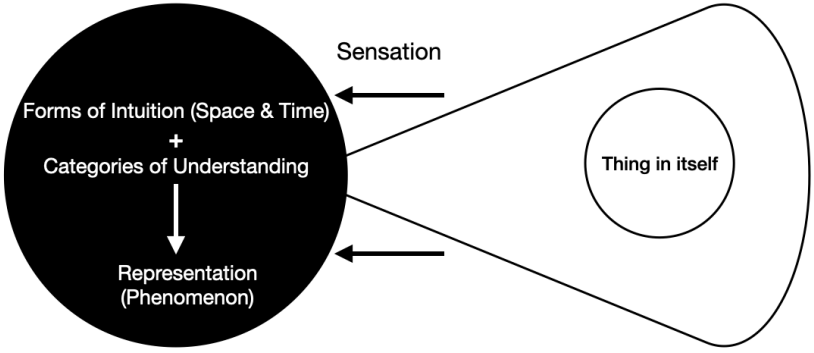
2. Finding an Enlightened Ground

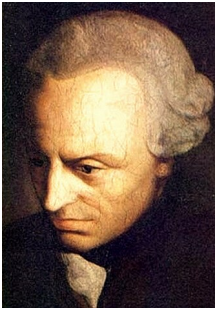
3. Breaking the Ground

4. Artificial Systems – A Paradox?

"[M]any people would accept that we do not really have knowledge of the world; we have knowledge only of our representations of the world. Yet we seem condemned by our constitution to treat these representations as if they were the world, for our everyday experience feels as if it were of a given and immediate world." – Francisco Varela

Transcendental Idealism





Kant searched for the conditions of making coherent experiences (Kant, 1781).

Kant argues that the mind does not just record sensory input (Hume)—it organizes it.

- Space and time are *a priori* **forms of intuition**: the mind's built-in ways of structuring all sensory data.
- The **categories of the understanding** (like causality, substance, unity, plurality) shape how we can think about what we perceive.

$\Rightarrow$ According to Kant, synthetic a priori knowledge is possible because the mind itself provides the a priori forms and concepts that structure all experience.

Kant thought that anything could be discussed and critiqued:

> *"For enlightenment of this kind, all that is needed is freedom. And the freedom in question is the most innocuous form of all—freedom to make public use of one's reason in all matters." – Kant*

But, of course, only reasonable people have this freedom! Woman were excluded.

This tradition of a belief in the **Voice of Reason** in **public discourse** is still alive (compare e. g. Habermas) but questioned by many e. g. Luhmann, Rorty.

That almost everyone can now "speak" has neither lead to an "enlightened society" (Kant) nor a "revolution of the oppressed" (Marx).

"Objectivity is the illusion that observations can be made without an observer [even Kant's observation of forms of intuition and categories of understanding]." – Heinz von Foerster

"I think we should not abandon [Kant's] idea of resistance, but we should relocate it into the system. It is the result of resolving an internal conflict—the result of the system's operations resisting the operations of the same system." – (Luhmann, 1995)

1. Why Luhmann's Systems Theory?

2. Finding an Enlightened Ground

3. Breaking the Ground

4. Artificial Systems – A Paradox?

Evolution of Cybernetics

Milestones

1. **Macy Conferences (1946–1953)**: Birth of 1st-order cybernetics: feedback, circular causality, information, early computing.
2. **1950s–1960s Expansion**: Control engineering, biology, anthropology/psychology; neural nets and symbolic AI diverge.
3. **2nd-Order Cybernetics (1960s–1970s)**: Observer included; von Foerster, Bateson, Maturana & Varela; BCL and ASC as hubs.
4. **Integration (1970s–1990s)**: Systems thinking, complexity, organizational cybernetics (Beer), constructivism; influence on therapy and design.
5. **2000s–Today**: Enactive/embodied cognition; complexity science; ML/AI, HCI, systems design and resilience.

Key figures: Wiener, McCulloch, Shannon, von Neumann, von Foerster, Bateson; Maturana & Varela; Beer, Mead.

First-order cybernetics

The study of the control of *observed* systems.

- **Focus:** How systems maintain stability and control via feedback.
- **Observer:** Treated as outside the system being studied.
- **Key question:**
  - How can we measure, predict, control, or regulate a system?
  - How does feedback produce order?
- **Domains:** engineering, biology, control theory, early AI.
- **Conceptual stance:** More aligned with a scientific realist perspective. The system has properties that the observer discovers.

"The creature that wins against its environment destroys itself."
– (Bateson, 1972)

Second-order cybernetics

The study of *observing* systems that cannot control their environment.

- **Focus:** How observers generate descriptions of systems—including how their own participation shapes what they observe.
- **Observer:** Included inside the system; observation is itself part of the dynamics.
- **Key question:**
  - How do observers construct reality?
  - How does cognition arise from self-organizing systems?
  - What are the implications of the observer affecting what is observed?
- **Domains:** cognition, social systems, communication, therapy, education, design.
- **Conceptual stance:** Constructivist, not necessarily realist. Reality is not "discovered" but co-constructed through interaction.

Second-order cybernetics

The study of *observing* systems that cannot control their environment.

- No universal a priori categories; cognitive structure arise from biological autopoiesis, learning, and interaction (strcutural coupling)
- Knowledge has viability rather than necessity: works for the system in its environment, but could be otherwise
- Limits of knowledge are operational, based on system organization

Comparison:

- **Kant:** The mind *must* structure the world in certain ways (necessary and universal).
- **2nd-order cybernetics:** Each observer *does* structure its world in certain ways (contingent, system-relativ).

What counts as "knowledge" is a product of a system's structure and history—not a transcendental a priori.

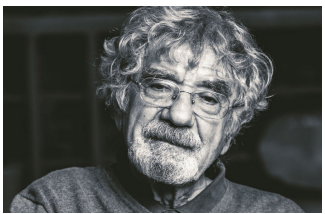
Information is no longer an object or a signal but a relational effect within a system capable of perceiving and responding to differences.

"Information is a difference that makes a difference." – (Bateson, 1972)

According to (Maturana & Varela, 1987), the nervous **system** is

1. energetically/materially open
2. semantically closed

meaning that **environmental perturbations** do not convey information; the system specifies its own changes.



"order from noise"

Key idea:

1. **Operational clousure:** A system can only operate through its own internal operations.
2. **Structural coupling:** Environment and organism co-evolve in interaction. The environment triggers changes, but the (nervous) system determines what those changes mean.
3. **No inputs as "information":** Sense organs do not transmit objective information; they trigger state changes that the nervous system interprets through its own dynamics (Maturana).
4. **Not isolation, but self-referential autonomy**: Operational closure does not mean isolation. The system can generate internal information by interpreting environmental events through its own code (Luhmann).

Operational Clousure

A system is operationally closed when it produces and links its own operations according to its own organization.

1. Psychic systems ("minds") *thinks* / *perceive* (Husserl & Luhmann)
2. social systems *communicate* (Luhmann) and
3. living systems realize metabolic processes (Maturana & Varela)

Operational closure enables systems to remain **autonomous**.

Structural Coupling

1. **Mutual Influence, No Instruction:** The environment does not specify what the system must do; it merely triggers changes. The system's own structure determines how it changes.
2. **History-Dependent:** Structural coupling is not a single interaction but a history of recurrent interactions that gradually align system and environment.
3. **Autonomy Preserved:** The system remains operationally closed: it only responds in ways permitted by its organization. Coupling never compromises autonomy.
4. **Co-Conditioning:** The organism changes with the environment, and the environment changes with the organism—a co-evolution of structures.
5. **Basis of Cognition and Meaning:** Cognition $\approx$ the activity of a structurally coupled, autopoietic system. What "counts as information" emerges from this coupling.

"We assume that systems exist" – (Luhmann, 1988)

System/Environment-Difference

1. A *system is* what it *does*: *operate*
2. It consists not of "things" but operations
3. The *world* is only accessible via the *environment*
4. A *system* is its difference to its environment (paradox?)
5. The environment is the exterior of the system
6. A system differentiates itself from its environment
7. All / some? systems *observe*
8. *Oberservation*, including self-observation via *re-entry* (Spencer-Brown, 1969) (paradox?) and higher order-observation means *indication* and *distinction* (Spencer-Brown, 1969)
9. *Oberservation* copies the system/environment distinction into the system
10. There is no observation from nowhere $\Rightarrow$ there is always a *blind spot*
11. *Resonance* is achieved whenever a signal in the environment affects the system's self-referential operation mode (Buchinger, 2012)

Autopoiesis

"We want to designate as autopoietic those systems that produce and reproduce the elements from which they are composed by means of the elements from which they are composed." – *Luhmann*

Autopoiesis

"Operation is the occurence of events, whoms reproduction performs the autopoiesis of the system, that is, the reproduction of the difference of system and environment." – (Luhmann, 2004)

$$
\begin{aligned}
& x^2 = 4 \\
\Rightarrow\ & x = 4/x \\
\Rightarrow\ & x = 4/(4/x) \\
\Rightarrow\ & x = 4/(4/(4/x)) \\
\Rightarrow\ & x = \ldots \\
\Rightarrow\ & x = 4/(4/(4/\ldots))
\end{aligned}
$$

A paradox outside of time becomes productive in time:

$$\begin{aligned} & i^2 = -1 \\ \Rightarrow\ & i = -1/i \\ \Rightarrow\ & i = -1/(-1/i) \\ \Rightarrow\ & i = \dots \\ \Rightarrow\ & i = -1/(-1/\dots) \end{aligned}$$

i enters itelf in its definition? If we set $i = 1$ we get

$$1 = -1/1 = -1.$$

If we set $i = -1$ we get

$$-1 = -1/(-1) = 1$$

both is clearly paradoxical. (Spencer-Brown, 1969) extends the Boolean algebra with the imaginary.

The Re-Entry: Embrace the Paradox

- I reflect on myself.
- Science communicates about its methods.
- The media reports on the media.
- ...

For Luhmann the characteristic of a modern society is its *functional differentiation*.

1. Why Luhmann's Systems Theory?

2. Finding an Enlightened Ground

3. Breaking the Ground

4. Artificial Systems – A Paradox?

Technology almost disappears (if it works) ("Handlichkeit" (Heidegger, 1977)).

- It is an evolutionary achievement that simplifies complexity.
- We (and social systems) can observe if it does *work* or it does *not work*.
- It reduces uncertainty in a way that results in *manageable ignorance*.

Some conceptialize technology as an autopoietic system, compare (Reichel, 2011; Watson & Romic, 2024; Lovasz, 2023).

Turing machines can be self-referential. They are

- **strictly/tightly coupled** with psychic systems, e. g. developers (and other Turing machines),
- *trivial machines* (von Foerster, 2003)
- **basal self-referential**, computations refer to other computations, e. g. a function calls itself.
- not *contingent* in themselves. It is precisely our effort to keep them deterministic such that they work.

Self-Learning Turing Machines

Self-learning Turing machines (ANNs) enable second-order observation of structures psychic systems cannot, by themselves, make sense of.

- **Loosely coupled** with psychic and social systems (and soon probably with other self-learning machines)
- They are **contingent** (despite being deterministic) because the data they are trained on is contingent.
- It is unlikely that they differentiate between *self-reference* and *other-reference*. They realize **no re-entry**, that is, they do not re-introduce the system/environment distinction into themselves.
- This would require that they are *meaning-based* systems (Buchinger, 2012).

Psychic and social systems co-evolved together which had led to a common achievement: meaning (Sinn).

Large language models couple with society via a common medium, that is, language!

- Can revisit, expand, or summarize their own previously generated text. But the text has to be provided.
- Using chain-of-thought they seem to trigger themselves: "Is this step correct." Computing the most plausible text which is often either a rejection or justification of what was written before.
- This is a form of rudimentary **processual self-referentiality** where plausibility assessment itself recursively shapes subsequent outputs.
- But again, there is **no re-entry** and it is difficult to argue that the architecture of LLMs support its emergence.

We agree with (Esposito, 2017, 2024) that LLMs participate in / shape communication and expand on this idea.

> *"However, even in the absence of a re-entry, interactions between language models and psychic systems nonetheless constitute a form of communication—despite one participant lacking any intrinsic capacity for sense making" – (Zönnchen, Dzhimova, & Socher, 2025)*

"Unlike classical machines, which are strictly coupled to psychic and social systems via formal languages, ANNs introduce a second, looser coupling through the 'digestion' of training data derived from memorized communcation of social systems. In this way, they parasitically 'absorb' social contingencies." – (Zönnchen et al., 2025)

Artificial System?

The term *artificial system* seems to be a contradiction. To become such a system, "AI's" have to

- generate their own operations,
- maintain their own organization,
- create a system-environment distinction (and re-enter), and
- reproduce themselves through their own output.

At the moment, we can only infer that they seem to tighten their coupling.

Artificial System?

"We" **observed** large language models as the following (Zönnchen et al., 2025):

- Language models function as cognitive systems in Luhmann's sense, in that they are structurally coupled to communication but possess only limited capacity, that is, limited processual and no reflective self-referentiality.
- Their outputs emerge through pattern selection based on internal probability distributions, yet without system-intrinsic sense-making.
- Their coupling with social and psychic systems is partly loose and partly strict: they depend on intransparent socially generated data and transparent optimization algorithms, their internal processes remain opaque.
- They extend beyond mere "parroting": instead of simply replicating existing text, they generate novel combinations of linguistic elements, which psychic and social systems subsequently interpret.
- They function simultaneously as both a medium and artificial communication partners, blurring the line between a tool and an autopoietic system.

Notes for AI Development

- There is no unbiased AI.
- Uniliteral control leads to catastrophe $\Rightarrow$ replace unilateral control with mutual coupling.
- Build in negative feedback loops.
- Do not only ask: "How can we make AI perform better" but "What **assumptions** about intelligence, knowledge, and value are built into our AI—and how can we learn from them?"
- Consider the human developer, data curator, and user as part of the feedback loop.
- Ask: "What is this system teaching us about ourselves?"
- Develop policy and regulation that learns and updates as AI systems evolve.

Disclaimer: Epistemic Humility

There is nothing outside of observations.

- Everything being thought, perceived and communicated were observations!
- But according to Luhmann, operational closure of a self-referential system cannot be directly observed from outside; it can only be inferred from its effects.
- Therefore, psychic systems or science will never be able to "prove" such closure if AI is not operating in the same medium (sense). **A reflexive grounding becomes impossible.**

Any questions?

References I

Bateson, G. (1972). *Steps to an ecology of mind*. San Francisco: Chandler Publishing Company.

Buchinger, E. (2012). Luhmann and the constructivist heritage: A critical reflection. *Constructivist Foundations*, *8*(1), 19–28.

Esposito, E. (2017). Artificial communication? The production of contingency by algorithms. *Zeitschrift für Soziologie*, *46*(4), 249–265. doi: 10.1515/zfsoz-2017-1014

Esposito, E. (2024). *Kommunikation mit unverständlichen Maschinen*. Residenz Verlag.

Gabriel, M. (2024). *Was ist eigentlich eine Künstliche intelligenz (ki)?* YouTube Video. Retrieved from `https://www.youtube.com/watch?v=jhJ1Wp4zemM` (Vortrag, veröffentlicht auf YouTube)

Gabriel, M. (2025). *Philosoph Markus Gabriel: KI, Gefühle, Macht.* YouTube Video. Retrieved from `https://www.youtube.com/watch?v=E1rfw2PR5MI` (Vortrag, veröffentlicht auf YouTube)

References II

Heidegger, M. (1977). *The question concerning technology, and other essays*. New York: Harper & Row.

Kant, I. (1781). *Kritik der reinen Vernunft*.

Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and Machines*, *17*(4), 391–444. doi: 10.1007/s11023-007-9079-x

Lovasz, A. (2023, jan). Niklas Luhmann and Jacques Ellul on the autonomy of technology. *Kybernetes*, *ahead-of-print*(ahead-of-print). Retrieved from `https://doi.org/10.1108/K-02-2023-0287` doi: 10.1108/K-02-2023-0287

Luhmann, N. (1988). Erkenntnis als Konstruktion. *Benteli*, 7–55.

Luhmann, N. (1992). *Die Wissenschaft der Gesellschaft*. Suhrkamp Verlag.

Luhmann, N. (1995). *Systems theory and postmodernism.* University lecture. Retrieved from `https://www.youtube.com/watch?v=wBBAkh52GhA`

Luhmann, N. (1997). *Die Kunst der Gesellschaft*. Frankfurt/M.: Suhrkamp Verlag.

References III

Luhmann, N. (1998). *Die Gesellschaft der Gesellschaft*. Suhrkamp Verlag.

Luhmann, N. (2004). *Die Realität der Massenmedien* (Vol. 3). VS Verlag für Sozialwissenschaften.

Maturana, H. R., & Varela, F. J. (1987). *Der Baum der Erkenntnis*. Scherz: Bern.

Reichel, A. (2011). Technology as system: Towards an autopoietic theory of technology,. *Interna-tional Journal of Innovation and Sustainable Development*, *5*.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. New York: Viking.

Spencer-Brown, G. (1969). *Laws of Form*. London: Allen and Unwin.

von Foerster, H. (2003). Cybernetics of cybernetics. In *Understanding understanding: Essays on cybernetics and cognition* (pp. 283–286). New York, NY: Springer New York. doi: 10.1007/0-387-21722-3_13

References IV

Watson, S., & Romic, J. (2024, jan). ChatGPT and the entangled evolution of society, education, and technology: A systems theory perspective. *European Educational Research Journal*. doi: 10.1177/14749041231221266

Zönnchen, B., Dzhimova, M., & Socher, G. (2025). From intelligence to autopoiesis: rethinking artificial intelligence through systems theory. *Frontiers in Communication*, *Volume 10 - 2025*. doi: 10.3389/fcomm.2025.1585321