

An argument for the impossibility of machine intelligence

Jobst Landgrebe Barry Smith

November 16, 2021
University at Buffalo
Department of Philosophy

jobstlan@buffalo.edu phismith@buffalo.edu

Abstract

Since the noun phrase ‘artificial intelligence’ (AI) was coined, it has been debated whether humans are able to create intelligence using technology. We shed new light on this question from the point of view of thermodynamics and mathematics. First, we define what it is to be an agent (device) that could be the bearer of AI. Then we show that the mainstream definitions of ‘intelligence’ proposed by Hutter and others and still accepted by the AI community are too weak even to capture what is involved when we ascribe intelligence to an insect. We then summarize the highly useful definition of basic (arthropod) intelligence proposed by Rodney Brooks, and we identify the properties that an AI agent would need to possess in order to be the bearer of intelligence by this definition. Finally, we show that, from the perspective of the disciplines needed to create such an agent, namely mathematics and physics, these properties are realizable by neither implicit nor explicit mathematical design nor by setting up an environment in which an AI could evolve spontaneously.

Keywords: Artificial Intelligence (AI), Complex Systems, Ergodicity, Phase Space, Context-dependence

1 Introduction

Since Turing [24] and others initiated Artificial Intelligence (AI) research as a discipline in its own right in the late 1940s, there has been a constant debate about whether AI can be achieved at all. Dreyfus [7] was one of the first to fundamentally question the feasibility of AI based on his Heideggerian view of the human mind. At the opposite extreme there are those – drawing on the incredible advances in AI in recent years in areas such as abstract games and machine translation – who have defended the idea that machines might be created with an intelligence vastly superior to that of our own species [4, 6].

The opinions expressed in this debate depend in every case on how ‘intelligence’ is to be defined. On the still standard view in the AI community, intelligence is defined as the ability of an agent to achieve a maximal reward sequence in a given environment at minimum computational cost [11]. We show in 2.1 that this and the other similar definitions proffered by the AI community are not only unable to capture human intelligence; they cannot capture even the intelligence manifested in insect behaviour. This is because they fail to model perception, motion (as a sensorimotor activity), and the non-Markovian, erratic nature of reward sequences.

A more substantial definition of intelligence is therefore required. To arrive at such a definition, to assess its implications for the issue of machine intelligence, we proceed as follows. First, we define what it is to be an agent that could be the bearer of AI. Then, we introduce a widely accepted and highly useful definition of basic (arthropod) intelligence. We then identify the properties that an AI agent would need to possess in order to be the bearer of intelligence by this definition. And finally we show that, from the perspective of the disciplines needed to create such an agent, namely mathematics and physics, these properties are unrealizable. This means that, no matter what mathematical models and engineering methods we deploy, we will still not be able to create a Brooks-intelligent agent of the needed sort, and thus also will not be able to create an intelligence comparable to that of human beings. Such methods may, of course, be developed in the future, but only if the nature of these disciplines is fundamentally transformed.

If we cannot engineer an intelligent agent, might it be possible to create the conditions under which an AI would emerge spontaneously – an idea advanced, for example, by Chalmers? In a coda, we show that this idea, too, is not supportable. We conclude that humans are unable to engineer AI agents, and that they are also unable to create a situation in which such agents could emerge spontaneously. Thus, there will be no intelligent machine in the sense of our proposed definition of intelligence, not even intelligence of the sort possessed by bees and other arthropods. It follows that higher intelligence (mammalian or human) is unfeasible also.

2 Engineering Intelligence?

2.1 Definition of ‘agent’ and of ‘Artificial Intelligence’

We define an agent as a man-made technical artefact that can perform actions in an open environment based on the events it senses in this environment. Thus it can perceive and move in ways that take account of what it perceives. For example, a self-driving car would be an agent, while an assembly-line robot or a confined self-driving train (such as are to be found in many airports) is not.

An intelligent agent must be able to interact with an open environment, namely the world that surrounds us. Any definition of intelligence that would not require this ability is insufficient. An agent unable to cope with natural environments would be less intelligent than the most primitive organisms. The AI-robotics pioneer Rodney Brooks recognized this already 30 years ago when he described his goal of building

robots that can ‘move around in dynamic environments, sensing the surroundings to a degree sufficient to achieve the necessary maintenance of life and reproduction.’ [5, p. 140]

The *utility-based* definitions proposed by mainstream AI authors [11, 18, 19, 22, 25] cannot meet this requirement. The utility function they provide¹:

$$V_{\mu}^{\pi} := \mathbf{E}\left(\sum_{i=1}^{\infty} r_i\right) \leq 1 \quad (1)$$

defines the utility V_{μ} of an agent π for an environment μ (where by environment is intended the corresponding minimal binary *description*), as the stochastic expectation \mathbf{E} over the infinitesimal series of the rewards r_i that the agent obtains as a result of its behaviour at each step.

The definition limits the maximum utility to 1 and implicitly ensures that simple environments which can be solved in a few steps will not receive a high utility score. Unfortunately it faces two main problems: (i) it is unable to take account of the active, iterative and directed, intentional nature of perception [8, 16].

Thus, for example, it cannot take account of what occurs when a predator actively observes the behaviour of its prey and follows it through changing environments. The static variable μ captures none of this. (ii) Its step-by-step, additive reward scheme cannot account for the sorts of reward patterns an agent can encounter in a real-world environment. Indeed, it is not possible to define on its basis even a reward sequence that mimics the reward pattern experienced by a natural intelligence such as a fruitfly, an aspect to which we return below.

On the basis of equation (1) Hutter (and the other authors listed) provide a definition of intelligence Υ as the sum of complexity-penalised² utilities an agent can achieve over a set U of environments defined for the agent:

$$\Upsilon(\pi) := \sum_{\mu \in U} 2^{-K(\mu)} V_{\mu}^{\pi} \quad (2)$$

where U is the set of environment descriptions with which the agent can cope, and V_{μ}^{π} is the utility achieved by the agent π for a given environment description $\mu \in U$. For AlphaZero, for example, U is the set of environments (board positions, opponent activities, etc.) occurring in games it can play. For a putative universal agent, of course, all possible environments would have to be taken into account [15, section 3.1]. Though the infinitesimal definition of utility in (1) and the penalisation of complexity in the definition of Υ provide a statistically robust measure of the kind of surrogate intelligence those working in the general artificial intelligence (AGI) field have decided

¹The definition displayed here is from Hutter [11], who refers to it as a definition of ‘universal intelligence’. The other cited definitions are isomorphic thereto. A more detailed statement of the mathematical and other problems associated with these definitions is provided in [15]

²The penalisation is achieved via the Kolmogorov complexity function K , which indicates the length of the shortest algorithm needed to computationally represent the environment description μ of the agent π .

to focus on, the definition is too weak to describe or specify the behaviour even of an arthropod. This is not only obvious from the issues already mentioned above, but also from the fact that algorithms which realise the reward-schemes proposed in (1) and (2) (for example, neural networks optimised with reinforcement learning) fail to display the type of generalisable adaptive behaviour to natural environments that arthropods are capable of, for example when ants or termites colonise a house.

2.2 What an AI agent needs

The definition given by Brooks, on the other hand, applies directly to natural environments. It is useful for our purposes, first, because it circumvents the underspecification of intelligence of the type shown in Hutter-style equations, which fails to model the way intelligent organisms interact with nature. But it is also useful because it avoids the thorny issues involved in providing a definition precisely of human-level intelligence. As Brooks saw, a machine at the intelligence level of an ant with a goal system that could be parameterised by humans would already be highly useful.

Brooks defines an AI agent, again, as an artefact that is able ‘to move around in dynamic environments, sensing the surroundings to a degree sufficient to achieve the necessary maintenance of life and reproduction’. The natural environment of a moth, for example, is a dynamic environment in the relevant sense, and so also are the hybrids of natural environment and man-made technology that we humans encounter on a city street or highway, in a shopping mall or nuclear reactor, or while flying a helicopter in a mountain valley. Such environments are ‘dominated’ by complex systems in the sense defined in thermodynamics [23], where a ‘system’ is defined as a totality of dynamically interrelated physical elements. To delimit a system is to select a level of granularity of its elements, from microphysical particles to entire galaxies, and to identify the types of interactions between such elements as the system exists through time.

All living organisms are complex systems [17, 20, 21], but so also are many inanimate systems, for example the Earth’s weather system, climate system, and geological water cycle [23]. Such systems share many features. Here we will examine only those features needed to show that an agent able to interact intelligently with such a system cannot – using the mathematics which exists today – be engineered. First, however, we need to contrast complex systems with what we shall call ‘logic systems’, which are systems modelled, designed and built using extended Newtonian mathematics, including the standard logical inference machinery.

2.2.1 Some properties of logic systems

All technical artefacts we build are logic systems in this sense, and in exercising their main functions their behaviour is deterministic.³

³Those which consume energy have a dynamic aspect which Prigogine called ‘dissipative’ (because the energy they consume is dissipated in the form of some kind of motion, radiation or heat). This is true, too, of your phone and your laptop, but it is an aspect that is in most situations not relevant for main their function or to our argument here.

There are some who hold that engineers often build technical artefacts not on the basis of mathematical models but rather by using trial and error, followed by physical testing of the results. There are indeed some who hold that something similar is the case when engineers build digital neural networks and other, similar stochastic artefacts. There may indeed be some engineered devices which we built in this way. But if they realize their intended functions then there is (or can be built) in every case a mathematical model which represents how they work with an approximation close to the measurement error.

In the case of neural networks the matter is still clearer. Neural networks are indeed built by a process of training, which means that they are fed with sample data selected as material for the algorithm to learn from. But neural networks are logic systems. They are engineered and trained to run on computers, which means that each one of them can of necessity be identified with some algorithm. This may be, highly complex – in the case of the GTP-3 algorithm, for example, it consists of some 175 billion parameters.⁴

There are three additional properties of logic systems of importance for our argument here:

1. Their phase space is fixed.
2. Their behaviour is ergodic with regard to their main functional properties.
3. Their behavior is to a large extent context-independent.

What does this mean?

Phase space A system can be mathematically characterised by variables describing the behaviour its elements. These variables are called ‘state variables’ in physics. The phase space of a system is the vector space whose basis is formed by the state variables relating to each element type of the system. This space has as many dimensions as the number of distinct state variables that relevant elements of the space can have. The variables in this space are determined by the way the system is modelled mathematically, and thus in any given case it consists of a finite set of measurable values. A fixed phase space may allow almost exact⁵ mathematical modelling of a system, for example using differential equations.

Ergodicity Ergodicity describes the stochastic behaviour of a system. A system is called ergodic if, over sufficiently long periods of time, the time in which a system element occupies any given region of the phase space is proportional to the volume of

⁴It is this merely algorithmic complexity which explains why the search for ‘explainable AI’ must often remain fruitless.

⁵‘Almost exact’ means that measurements performed on a system will deviate only from the model in the order of magnitude of the measurement error. For example, the basic Newtonian model of the earth’s orbit around the sun based on the gravitational interaction of these two bodies is almost exact, but does not take into account the influence of other bodies, which lead to slight periodic deviations of the earth from its orbit.

this region. Access to the regions of the phase space is thus equiprobable over time. For example, a steam engine, which is a logic system when account is taken only of its behaviour in exercising its main function,⁶ the elements (for example the pistons in the engine's cylinders) access all regions of the phase space in an equiprobable manner. This behaviour of the engine is deterministic and the speed of a locomotive driven by the engine can therefore be regulated via the pressure in its boiler.

Context-independence A logic system is self-contained. Its elements interact in the same manner independently of the system's environment as long as this environment does not alter the phase space of the system. For example, the solar system as a gravitational system is a logic system, and its behaviour would not be altered if it was moved some light weeks closer to the binary star system Alpha Centauri (which is 4.37 light years from the sun).⁷ In the same way, a motor vehicle is a logic system that can be used on any terrain for which it was built, no matter where it is – as long as there is gravitation, an approximatively even ground, a slope not surpassing a certain angle, and oxygen in the air for its combustion engine.

Note that the goal of AI research, in our terms, is to build a logic system that is at least Brooks-intelligent.

2.2.2 Some properties of complex systems

The systems that make up our environment and interact with each other are not logic systems but complex systems. Our world is shaped by complex systems, some of them inanimate, such as the tectonic movements caused by the Earth's seismic system; and some of them animate, in particular all the behaviour of humans and other animals. As we learn from we standard mathematical theory of complex systems [23], all such systems, including the systems of complex systems resulting from their interaction,

1. have a variable phase space,
2. are non-ergodic, and
3. are context-dependent.

(1.) A **variable phase space** means that the variables which define the elements of a complex system can change over time. New elements and new associated variables can arise, and others can disappear. Existing elements can obtain new variables, or in other words new dimensions along which their behaviour may vary, as occurs for instance when a child learns to walk, or to swim, or to play tag. The way the elements of the system interact can change. New types of interaction can arise. All this applies, for example, to the way people evolve over their lives, to the interactions of animals in groups, or to the way conversations (and indeed entire languages) evolve over time.

⁶The steam engine is a complex system when account is taken of the energy dissipation occurring in its steam boiler.

⁷Things would be different if the sun came so close to Alpha Centauri that they would gravitationally interact in a non-negligible way.

(2.) A **non-ergodic** system produces erratic distributions of its elements. No matter how long the system is observed, no laws can be deduced from observing its elements. A simple example is a group of children in a school playground. Though patterns emerge, they can change at any time and for any one of an open-ended number of reasons.

(3.) A **context-dependent** system changes its elements and the way they interact depending on its environment. Examples are (i) the total dependence of those species which are specialized to specific environments upon those environments (which means that they perish in other environments), and (ii) the adaptation of certain species to multiple environments which lets them change their behaviour when they switch from one environment to another.

2.3 Why we cannot model, design and build an agent that is Brooks-intelligent

Each AI agent, by our definition of ‘agent’, will have to cope with a complex-system-generated environment. An AI agent is a technical entity and needs to be modelled, designed and built (engineered). Modelling such an agent means to define its functions and the way it is supposed to interact with its environment. Designing it means to specify how these functions are to be realised physically. Both steps require mathematical models.

There are two fundamental types of mathematical model: deterministic and stochastic (along with hybrids, usually obtained by composition). Any mathematical model requires a vector space, often a coordinate space such as the Cartesian space \mathbb{R}^2 or the Euclidean space \mathbb{R}^3 over an algebraic space such as that formed by the real numbers \mathbb{R} . Such a space is needed to define the variables used in the mathematical equations which are supposed to model the corresponding real system.

Variable phase space But with the *changing variables and interactions* that we find in complex systems, there is *no coordinate space* over which models can be defined. Any model is defined for a specific vector space, and is invalid if the reality targeted by the model differs from the vector space for which it was defined. The more it differs, the stronger is the deviation, and the less accurate the model becomes. With complex systems, the deviation can become huge, and this can happen suddenly, when new elements evolve.⁸

Examples of such deviation, which arise due to the massive discrepancy between model and system, are reflected in the predictions generated in a weather forecast (which are never accurate over longer periods), climate models (which are never accurate) or epidemiological models (which are in extreme situations only approximatively accurate). All these deviations become larger as time passes, because the effects which prevent exact models for complex systems gain importance exponentially [23].

⁸This is true even though partial aspects of the system can be modelled approximatively. For example, we have an approximative, descriptive understanding of the physiology of blood pressure. Yet we do not understand essential hypertension and have no causal way to treat it once it emerges (unpredictably)[12].

Non-ergodic process Yet more obstacles to modelling are created where we are dealing with *non-ergodic processes*, which produce events in which we cannot identify *any* law-like pattern that could be modeled mathematically. Of course, there are many patterns in complex systems which can be modelled. These include for example the different sorts of cyclic behaviour exhibited by all animals along multiple dimensions (respiration, heart beat, wake-sleep-cycle, feed-rest-hunt cycle, cycles of sexual behaviour, etc.). But these cover only partial processes and cannot be described using differential equations or other exact modelling approaches. For example, we cannot even exactly model classical conditioning in a way that would be valid across multiple laboratories experimenting on harnessed arthropods (honey bees) even where the exact same experimental setup is being applied [2].

Importantly, the traces of non-ergodic processes (which means the series of data which such processes generate) provide no adequate target spaces for stochastic sampling, because the latter requires in every case that samples are produced from a distribution belonging to the exponential family or to a non-parametric distribution that can be reliably estimated [13]. The samples drawn from such traces are never representative of the process behaviour due to the non-ergodic character of the process. There is no distribution family to sample from. This systematically prevents stochastic modelling of such processes. It is also the reason why reward traces like those described in equation (1) cannot be provided for the training of artificial agents where interactions with complex systems are involved.

There is no way to set up such a trace because there is no stochastic process available to create it. A deterministic model would not work either because such a model, too, cannot produce the pattern of rewards produced by a complex system.

This is the reason why, for example, machines cannot learn to conduct real conversations using machine learning – which is a method to obtain mathematical models implicitly using optimization methods applied to training data [1, 9, 10]. The training data obtained from recorded conversations are never adequate as material to train a model because the conversations are conducted by complex systems. If such data are used to train a stochastic model, the model will always be fitted to non-representative data. It will therefore fail, if not in the next conversation, then in some conversation in the near future, and not in the way that humans fail regularly in conversations, but rather in a way that betrays that the model is not able at all to meet basic human expectations for a conversation with another human being [14].⁹

Fully deterministic conversation agents are improving every day even as we speak. But more than fifty years after the first chatbots appeared, they seem still not to have reached the level where they can successively and smoothly book a room or a flight without requiring extra effort on the part of the human being involved. The goal of a conversation agent that can make the sorts of spontaneous responses over a longer period that are characteristic of human beings is unreachable.

⁹Willingness to chat with a bot for entertainment purposes, a habit quite popular in some cultures, is a different matter.

Context-dependence This property of complex systems has the consequence that the system will use a different phase space following different principles depending on the context in which it is situated. A housefly cannot adapt to a window pane. Its behaviour becomes unpredictable and erratic when it gets caught behind a window. On the other hand, those animals that can adapt to new types of situations show a completely erratic pattern during the process of adaptation. Because, in natural environments (as in financial markets), the context can radically change at any time, context dependence makes the behaviour of complex systems in such environments unpredictable. Of course, humans have developed strategies to make each other's behaviour more predictable in order to enable life in communities (based on mutual trust) and societies (based on social norms). But the omnipresence of conflicts (private and public, peacefully resolved or leading to violence) shows that predictability is still very low, since the interaction of complex systems is creating novel, unexpected situations all the time.

2.3.1 The self-driving car amidst complex systems

To summarise, an artificial agent, to emulate human-level intelligence, would have to cope with an environment arising from the interactions of animate and inanimate complex systems. But we have no way of modelling such interaction mathematically. Neither explicitly, using manually crafted models, nor implicitly, using stochastic models.¹⁰

When an AI agent is interacting with simple systems, for example an agent built for purposes of missile defense, the system works by modelling the motion of the incoming missiles, which are simple systems, using Newton's Laws. The phase space of the missiles is fixed, their behaviour is Markovian, their trajectory is ergodic.

When, in contrast, the AI agent is in the situation of a self-driving car (agent) moving through the dense traffic in downtown San Francisco or through the mountain passes surrounding Lake Tahoe two hundred miles to the North, then in both situations, the variables it may be confronted with variables which are permanently changing. Although for the untrained eye of a non-mathematician, the car simply moves in Euclidean space over time, when viewed through the eyes of an engineer, it faces a huge phase space determined by the various entities crossing its path. Because both animate and inanimate systems often show erratic behaviour, new phase space dimensions not present in the training data can arise any time, for example if a flock of birds flies over a road at low altitude while a child is running onto the road to catch its ball; or a demented pedestrian stumbles across the road in a dense fog massively diffracting the car's lights. The total situation the car is facing is determined by a complex system of complex systems whose behaviour is at every level non-ergodic. We are unable to sample adequate training distributions from the traces created by such processes, and unable also to define an adequate reward-path to train a utility-based agent.

¹⁰However, we can model partial aspects very successfully using AI [14].

2.4 Coda: Why AI agents cannot arise spontaneously

Could we create a system in which intelligence could evolve spontaneously, as proposed by Chalmers [6, p. 16]? His argument runs as follows:

1. Evolution produced human-level intelligence.
2. If evolution produced human-level intelligence, then we can produce AI (before long).

-
3. Absent defeaters, there will be AI (before long).

The first premise is of course true. But the second is false, because it implies an equivalence between evolution and human activity, an equivalence which does not obtain. Intelligence arose by natural selection through interactions of our ancestors with the inanimate environment and with competing organisms. There was no *plan* to yield intelligence on the basis of which these processes occurred, but rather only the *drive* to survive and reproduce that is built into every living organism. This drive led to the rise of intelligence through an immense series of processes, almost all of which occurred by chance.

But we neither know how to engineer the drive that is built into all animate complex systems, nor do we know how to mimic evolutionary pressure, which we do not understand and cannot model (outside highly artificial conditions such as a Petri dish). In fact, if we already knew how to emulate evolution, we would in any case not need to do this in order to create intelligent life, because the complexity level of intelligent life is lower than that of evolution. This means that emulating intelligence would be much easier than emulating evolution *en bloc*.

Chalmers is, therefore, wrong. We cannot engineer the conditions for a spontaneous evolution of intelligence.

3 Conclusions

As we saw earlier, a machine at the intelligence level of an ant with a goal system that could be parameterised by humans would be highly useful. But if we cannot create artificial intelligence even at this level, this means that we cannot create any artificial agent that deserves to be called ‘intelligent’, since Brooks’ definition targets intelligence of a very much weaker form.

This has two consequences:

1. We have no reason to fear artificial superintelligence à la Bostrom [4].
2. We can still engineer very useful ‘narrow AI’.

To the first point. Given that we have chosen a very weak definition of intelligence to prove that even this cannot be engineered, we are also on the safe side regarding the question whether humans might one day be overrun by an AI of any sort. For if we cannot engineer an arthropod level AI, then we cannot engineer, either, an AI that could initiate a process that would lead to AIs that could surpass human intelligence.

To the second point. The advances in applied statistical learning since the late 1990s have been very impressive. We can now solve many specific problems identified by humans using machine learning or a combination of machine learning with other algorithms. We will see many very useful applications of this technology over the next decades, not only in the design and supervision of ever more sophisticated logic system devices, but also in machine interactions with animate complex systems. However, these solutions will be architected by humans and dedicated to solving special problems, either using explicitly designed models or using optimisation procedures applied to training data. The utility of the latter depend always on the quality and scope of the training data, trained functionals or operators cannot cope with data points which do not or only rarely occur in the training set. They do not generalise, despite recent progress with stochastic ‘foundation models’ [3]. Therefore, they will always be restricted to a small scope of narrow AI, even if they can be adjusted to various such scopes with little training effort. But as the scope increases, the non-ergodic nature of the sampling distribution becomes tangible in the training data and the models become overfitted to samples that differ from reality.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] Dimitri P Bertsekas. *Nonlinear Programming*. Belmont, MA: Athena Scientific, 2016.
- [2] ME Bitterman. “Classical Conditioning since Pavlov”. In: *Review of General Psychology* 10.4 (2006), pp. 365–376.
- [3] Rishi Bommasani et al. *On the Opportunities and Risks of Foundation Models*. 2021. arXiv: [abs/2108.07258](https://arxiv.org/abs/2108.07258) [cs.LG].
- [4] Nick Bostrom. *Superintelligence. Paths, Dangers, Strategies*. London: Oxford University Press, 2003.
- [5] Rodney A Brooks. “Intelligence without representation”. In: *Artificial intelligence* 47.1-3 (1991), pp. 139–159.
- [6] David J. Chalmers. “The Singularity: A Philosophical Analysis”. In: *Journal of Consciousness Studies* 17 (2010), pp. 7–65.

- [7] Hubert L Dreyfus. *What Computers Can't Do: The Limits of Artificial Intelligence*. Cambridge, MA: MIT Press, 1972.
- [8] James J. Gibson. *An Ecological Theory of Perception (1979)*. Boston, MA: Houghton Mifflin, 2015.
- [9] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. Cambridge, MA: MIT Press, 2016.
- [10] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. 2nd ed. Berlin: Springer, 2008.
- [11] Marcus Hutter. “Universal Algorithmic Intelligence: A Mathematical Top-Down Approach”. In: *Artificial General Intelligence*. Ed. by Ben Goertzel and Cassio Pennachin. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 227–290.
- [12] Jens Jordan, Christine Kurschat, and Hannes Reuter. “Arterial hypertension: diagnosis and treatment”. In: *Deutsches Ärzteblatt International* 115.33-34 (2018), p. 557.
- [13] Achim Klenke. *Probability Theory: A Comprehensive Course*. 2nd ed. New York, Berlin: Springer, 2013.
- [14] Jobst Landgrebe and Barry Smith. “Making AI Meaningful Again”. In: *Synthese* 198 (2021), pp. 2061–2081. DOI: <https://doi.org/10.1007/s11229-019-02192-y>.
- [15] Jobst Landgrebe and Barry Smith. *Press zero for a human operator. Why there will be no AI*. Manuscript under review.
- [16] Maurice Merleau-Ponty. *Phenomenology of Perception (1945)*. London: Routledge, 2012.
- [17] Peter T Mora. “Urge and Molecular Biology”. In: *Nature* 199.4890 (1963), pp. 212–219.
- [18] Luke Muehlhauser and Louie Helm. “The Singularity and Machine Ethics”. In: *Singularity hypotheses: A Scientific and Philosophical Assessment*. Ed. by Amnon H. Eden and James H. Moor. Dordrecht: Springer, 2012, pp. 101–125.
- [19] Cassio Pennachin and Ben Goertzel. “Contemporary Approaches to Artificial General Intelligence”. In: *Artificial General Intelligence*. Ed. by Ben Goertzel and Cassio Pennachin. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 1–30.
- [20] Ilya Prigogine. *Introduction to Thermodynamics of Irreversible Processes*. Charles C. Thomas Publishers, 1955.
- [21] Ilya Prigogine and René Lefever. “Theory of Dissipative Structures”. In: *Synergetics*. Ed. by H. Haken. Wiesbaden: Vieweg+Teubner Verlag, 1973, pp. 124–135.
- [22] Jürgen Schmidhuber. “Gödel Machines: Fully Self-Referential Optimal Universal Self-improvers”. In: *Artificial General Intelligence*. Ed. by Ben Goertzel and Cassio Pennachin. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 199–226.
- [23] Stefan Thurner, Peter Klimek, and Rudolf Hanel. *Introduction to the Theory of Complex Systems*. Oxford: Oxford University Press, 2018.

- [24] Alan Turing. *Can Digital Computers Think?* 1951. URL: <http://www.turingarchive.org/browse.php>
- [25] Eliezer Yudkowsky. *Coherent Extrapolated Volition*. San Francisco: The Singularity Institute, 2001.