A Comprehensive Examination of Universal Artificial Intelligence: Unraveling the Concept of Intelligence

Paul Helstab PAUL.HELSTAB@UNI-ULM.DE
1169401

Abstract

This paper aims to comprehensively examine the concept of intelligence through the lens of Artificial General Intelligence (AGI), particularly focusing on Marcus Hutter's concept and formal definition of Universal Artificial Intelligence (UAI) and the AIXI model. It delves into the core principles of universal intelligence, laying out its theoretical underpinnings, exploring potential limitations, and considering its impact in the broader context of general intelligence research. The paper culminates in an analytical discussion, highlighting an intriguing parallel between intelligence and compression, while critically evaluating this perspective in relation to the central research questions.

1. Introduction

Marcus Hutter's definition of intelligence, which posits intelligence as "an agent's ability to achieve goals in a wide range of environments", is particularly significant as it provides a universal, non-anthropocentric perspective (Hutter, 2005). This viewpoint is essential when considering Artificial General Intelligence (AGI), as it broadens our understanding of intelligence beyond just human intelligence. It also avoids the need to list specific elements of intelligence, which can be limiting and contentious given the abstract nature of the concept (Legg & Hutter, 2007). This paper begins is exploring Marcus Hutter's definition of intelligence and the term Universal Artificial Intelligence (UAI). Subsequently, the paper delves into these definitions, emphasizing potential limitations, methods of testing, and the implications of this concept of intelligence. This analytical discourse is organized to address the following three research questions:

1. What are the practical and theoretical limitations of Hutter's universal intelligence agent, AIXI, in terms of its computability and its ability to accurately represent the concept of general intelligence?

2. What are the limitations of testing the intelligence of universal agents according to the definition by Hutter and Legg, particularly in relation to the concept of general intelligence?

3. What are the potential implications of Hutter's concept of Universal Artificial Intelligence for the field of AGI research and development, particularly in aspect to Research Questions 1 and 2?

By systematically addressing these research questions, the paper aims to provide a comprehensive understanding of Hutter's concept of universal intelligence, unraveling the complexity of defining intelligence in the development of a potential AGI.

2. Definitions

Efforts to define and understand intelligence have led to various perspectives throughout the years. One particularly comprehensive and mathematically grounded definition is provided by Marcus Hutter and Shane Legg.

Definition 1 (Intelligence (Hutter, 2005))**.** *The measure of an agent's ability to solve an unknown goal in a wide range of unknown environments.*

Following this definition of intelligence, the next logical step is to delve into the concept of universal intelligence. This involves the theorethical development of an artificial system, such as AIXI, that is capable of demonstrating intelligence across a broad spectrum of unknown environments, thereby embodying the essence of Hutter and Legg's definition (Hutter, 2005). Defining and mathematically formalizing intelligence presents a complex problem from various perspectives. Intelligence can originate from diverse sources such as creativity, pattern recognition, musicality, problem-solving, learning, optimization, survival, and many more (Hutter, 2005).

2.1 Universal Predictor

The definition of universal intelligence is based on defining an universal predictor, which is essentially an enhanced version of the Bayesian sequence prediction model tailored for inductive reasoning (Hutter, 2006, 2007a). This model aims to predict future events or data based on past observations, using a loss function to measure the difference between the predicted and actual outcomes, and maximizing the posterior probability (Hutter, 2005). The expansion, as envisioned by Ray Solomonoff, leads to a so called universal sequence predictor not limited to a specific hypothesis class or prior distributions (Hutter, 2007a). Instead, it is capable of learning and making predictions from a universal class of environments, making it more flexible and general (Hutter, 2005).

2.1.1 Solomonoff's theory of inductive inference

Solomonoff combined the ancient Epicurus principle of multiple explanations with the Occam's Razor principle and the Bayesian inference framework to form a universal theory of inductive inference, also known as Solomonoff induction (Rathmanner & Hutter, 2011).

Definition 2 (Epicurus' principle of multiple explanations)**.** *(Hutter, 2005) If more than one theory is consistent with the observations, keep all theories.*

Definition 3 (Occam's razor)**.** *(Hutter, 2005) Entities should not be multiplied beyond necessity.*

Solomonoff induction is a method of prediction that is based on all possible hypotheses, weighted by their simplicity, as measured by the length of their description. In essence, it is a formalization of Occam's Razor, favoring simpler theories to complex ones (Hutter, 2005). This theoretical framework forms the basic idea for the definition of Marcus Hutter's universal predictor (Hutter, 2007a).

2.1.2 Information Theory and Kolmogorov Complexity

Shannon entropy is a fundamental concept from information theory that measures the average rate at which information is produced by a stochastic source of data (Hutter, 2005). It quantifies the amount of uncertainty or randomness in a data source. For instance, in the context of a library, a book with high Shannon entropy would contain a sequence of words that is very unpredictable (Hutter, 2005, 2020b).

Independently, Ray Solomonoff and Andrey Nikolaevich Kolmogorov published the concept of Kolmogorov complexity, which is defined as the length of the shortest computer program that can produce a given object as output (Hutter, 2005).

Definition 4. *Formally, Kolmogorov Complexity for an object x is defined with respect to a universal Turing machine U as follows:*

$$K(x) := \min_p \{\ell(p) : U(p) = x\}, \quad K(x|y) := \min_p \{\ell(p) : U(\langle y' p \rangle) = x\}$$

where $\ell(p)$ denotes the length of program p, and $K(x|y)$ represents the conditional Kolmogorov complexity of x given y (Hutter, 2007b)

This measure shares many properties with Shannon entropy and also extends the Information Theory by incorporating a computational aspect through an algorithmic solution (Hutter, 2005). Using the library analogy, the Kolmogorov complexity of a book is akin to the length of the shortest possible set of instructions needed to recreate the book from scratch (Hutter, 2005, 2020b). A book with a highly predictable pattern, such as one where every page is filled with the letter "A", would have low Kolmogorov complexity because the instructions to recreate it could be very brief, like "write 'A' on every page until the book is complete." Conversely, a complex novel would have high Kolmogorov complexity because the shortest program to reproduce it would be as lengthy as the novel itself (Hutter, 2005, 2020b). Marcus Hutter uses this concept as a measure of the computational resources required to specify an object within AIXI, which is explained in section 2.1.3.

2.1.3 AIξ Model and the AIXI Agent

Hutter's AIξ model is a theoretical mathematical formalism that aims to define a universal agent capable of learning and succeeding in any task within an unknown environment (Hutter, 2000). This model combines Ray Solomonoff's inductive inference and Kolmogorov complexity to create a universal predictor (Hutter, 2005). The AIξ Model leverages these principles to make predictions and favor simpler theories, respectively (Hutter, 2005).

Definition 5 (AIXI)**.** *(Hutter, 2005) The AIXI model, a_k and a_m denote the actions taken by the agent at times k and m respectively, while o_k and o_m are the corresponding observations, and r_k and r_m are the rewards received. The model employs a program q that, when executed on a universal Turing machine U, generates these observations and rewards based on the agent's actions. The Solomonoff prior probability of program q, represented as $2^{-\ell(q)}$, embodies the principle of Occam's razor, favoring shorter (simpler) programs.*

$$\textit{AIXI} \quad a_k := \arg\max_{a_k} \sum_{o_k r_k} \ldots \max_{a_m} \sum_{o_m r_m} [r_k + \cdots + r_m] \sum_{q:U(q,a_1\ldots a_m)=o_1 r_1 \ldots o_m r_m} 2^{-\ell(q)}$$

3. Analysis of the concept of universal intelligence

The subsequent chapter critically discusses how Hutter's definitions of universal intelligence relate to the testing and measurement of intelligence, the drawbacks, and implications concerning the concept of intelligence.

3.1 Drawbacks of universal intelligence

The ensuing discourse aims to critically examine the drawbacks associated with the concept of intelligence as delineated by the universal intelligence framework proposed by Marcus Hutter.

3.1.1 Abstracting intelligence as compression

In terms of intelligence, Hutter argues that the AIXI model represents the most intelligent unbiased agent possible (Hutter, 2005). It learns and adapts through reinforcement, aiming to maximize the total future reward provided by the environment (Hutter, 2005). This model is universal in the sense that it considers all possible environments and computable hypotheses, evaluating the rewards each program generates and selecting the action that has the highest expected total reward (Hutter, 2005).

The two theorems defined in sections 2.1.1 and 2.1.2 have two essential implications for this definition of intelligence. Firstly, Kolmogorov Complexity formalizes Occam's Razor by opting for the simplest solution to a problem and suggests that intelligence, in a broader sense, is merely a form of compression. Thus, AIXI is an intelligent system that can reduce the Kolmogorov complexity of data (i.e., find a shorter program that generates the data) and thereby reduce its Shannon entropy (i.e., make the data more predictable) (Hutter, 2005). Therefore, the connection between Kolmogorov complexity and Shannon entropy serves to substantiate further the assertion that intelligence is a form of compression, as elucidated in section 2.1.2. This perspective on intelligence is somewhat supported by the establishment of the Hutter Prize, an initiative that offers financial rewards for developing a compression algorithm capable of condensing a 100-million-character English Wikipedia article. The current prize fund stands at 500,000 euros, highlighting the significant investment in this research area (Hutter, 2020a).

While this definition of intelligence is comprehensive in its mathematical and theoretical approach, it can be criticized for its lack of consideration for aspects such as creativity, free will, emotions, and other subjective human experiences (Hutter, 2005). Hutter argues that these properties are only relevant if they have a measurable effect on performance in a well-defined environment (Hutter, 2005; Legg & Hutter, 2007). By this abstraction, viewing intelligence as the ability to find the shortest program and make predictions, an intelligent system would be one that can effectively compress data by finding these patterns. However, this is essentially what a good compression algorithm does: it finds patterns in the data and uses them to encode the data in a more compact form. This approach, which defines intelligence in a specific way, may fail to yield a new Beethoven symphony, as it tends to overlook for example musical intelligence, an important domain of intelligence. Hutter notes on this matter that AIXI must exhibit high creativity, but the resulting reward needs to be formally defined and proven (Hutter, 2012b).

3.1.2 Computability and the time is infinite problem

The most significant criticism of Hutter's AIXI model concerns its practicability, due to the calculation of the Kolmogorov complexity, as introduced in section 2.1.2. For every non-trivial (i.e., complex or non-obvious) object, this calculation is computationally intractable, meaning there is no efficient algorithm to determine it (Hutter, 2005). This difficulty arises because finding the shortest program that can produce a given output is a problem similar to the halting problem in computer science, which is known to be undecidable (Hutter, 2005). In this context, there is no efficient algorithm to solve it. Efficiency is typically defined in terms of polynomial time, meaning that the time it takes to solve the problem grows polynomially with the size of the input. If the best possible algorithm for a problem requires time that grows faster than any polynomial (for example, exponential time), then the problem is considered intractable (Sedgewick & Wayne, 2016; Hutter, 2005). Due to this issue, experts have raised concerns. For instance, Katayama argues, "AIXI does not adequately model the real world because it permits incomputable agents." (Katayama, 2019).

A solution to this issue is approximating the AIXI model (Katayama, 2019). Examples of approximations of the AIXI model include Hutter's AIXI-tl model (Hutter, 2005), Veness's MC-AIXI (Veness, Ng, Hutter, Uther, & Silver, 2010), and more recent approaches like Katayama's UCAI (Katayama, 2019).

AIXI-tl, a theorethical computable version of AIXI that only considers hypotheses of length l that run for less than time t (Hutter, 2005). This makes it a time and space-limited version of AIXI, which performs at least as well as the provably best time t and space l-limited agent (Hutter, 2005). However, AIXI-tl is not practically computable due to its size and complexity (Hutter, 2005). This theoretical version of AIXI is computable but limited in the environmental description length and the computation time per time step on a sequential Turing machine (Katayama, 2019; Hutter, 2005).

MC-AIXI, short for Monte Carlo AIXI, employs a Monte Carlo method to approximate the intractable aspects of the AIXI model (Veness et al., 2010). This approach represents an approximation of the AIXI model, tailored to a limited class of environments (Veness et al., 2010). Offering a more feasible solution, MC-AIXI is a computationally practical version of AIXI. It is crucial to recognize that while MC-AIXI is a computable version, it essentially serves as an approximation of the original AIXI model, due to the complex and infinite nature of the operations involved in AIXI (Veness et al., 2010).

Although these approximations can be theoretically computed and may perform well in certain scenarios (for example playing chess, go, tic tac toe or pacman), they do not fully capture the power of all properties defined by the original AIXI model (Veness et al., 2010; Hutter, 2005). Furthermore, this suggests that universal intelligence, in this context, is closely related or constrained by computational and physical resources, a problem also mentioned in Chalmers' 'The Singularity: A Philosophical Analysis' (Chalmers, 2010).

3.2 Testing universal agent's intelligence

The question of defining intelligence is closely related to how to test or measure the intelligence of a system, machine, human, or animal. This becomes especially complex when finding a universal test for all domains and environments. In his dissertation, Shane Legg offers a comprehensive exploration of various problems and corresponding solutions within this topic, which are discussed in this subsection.

3.2.1 Human intelligence tests

Assessing intelligence rapidly becomes a multifaceted endeavor, particularly when the scope is restricted to human intelligence and excludes broader, domain-general evaluations. Legg highlighted the complexities inherent in what are termed 'static tests' (Legg & Hutter, 2007). These tests, such as traditional IQ assessments, consist of pre-established questions that remain consistent in difficulty, irrespective of the test-taker's responses. Typically, the outcomes of these tests are benchmarked against a Gaussian normal distribution. An instructive example provided by Legg (Legg & Hutter, 2007), which is also cited by Jürgen Schmidhuber in a talk (Schmidhuber, 2016), illustrates the limitations of static intelligence tests measuring human intelligence. These tests consist of predefined questions with fixed difficulty levels and typically benchmark results against a Gaussian normal distribution. In one such test, a number sequence prediction task, participants are presented with the sequence 2, 4, 6, 8..., and the expected response is 10, following the straightforward linear pattern $2n$ (where n represents the position in the sequence, starting with 1) (Legg & Hutter, 2007). However, a response of 34, justified through the polynomial $n^4 - 10n^3 + 35n^2 - 48n + 24$, might demonstrate a higher level of intelligence (Legg & Hutter, 2007). In the context of these static tests, this more complex and equally valid answer could be mistakenly marked as incorrect, highlighting the challenges in capturing the nuances of human intelligence through such assessments. When considering that the essence behind the definition of universal intelligence is the adoption of the simplest solution by Occam's Razor (see section 2.1.1), a contradiction in the definition itself also becomes apparent here (Hernández-Orallo & Dowe, 2010).

3.2.2 Dynamic vs. static tests

Dynamic tests, in contrast to static tests, involve a greater degree of interaction between the test subject and the tester (Legg & Hutter, 2007). The tester presents the individual with a series of problems and provides feedback after each attempt, allowing the individual to adapt their behavior. This makes dynamic testing more costly to perform and increases the danger of tester bias (Legg & Hutter, 2007). Legg argues that this adaptive testing process would make dynamic testing more feasible and less biased when evaluating machine intelligence, which is a key aspect of measuring universal intelligence (Legg & Hutter, 2007).

3.2.3 Measuring universal agents intelligence

The concept of universal agents introduces the idea of a more general test of intelligence that could be applied to a variety of entities, from humans to machines to animals (Legg & Hutter, 2007). Such a test would ideally be dynamic, unbiased, fundamental, formal, objective, and fully defined (Hutter, 2005; Legg & Hutter, 2007). It would take into account the ability to learn and adapt over time, and would not be biased towards any particular dimension such as for example race, gender, culture or social class (Legg & Hutter, 2007).

In his paper, Legg posits that the testing of universal intelligence is more comparable to a definition than an actual test. This perspective is also influenced by the fact that calculating Kolmogorov complexity is computationally intractable, as discussed in section 3.1.2.

Furthermore, Legg and Hutter's definition of intelligence, which is based on the ability to achieve goals in a wide range of environments, could potentially be biased towards machine intelligence(Legg & Hutter, 2007). Machines are typically better at adapting to a wide range of structured environments and in learning through idea of compressing data (see 3.1.1).

The idea of testing intelligence through a universal distribution of environments, as proposed by José Hernández-Orallo, is an attempt to overcome some of the limitations of static and dynamic tests (Hernández-Orallo & Dowe, 2010). This approach involves testing an agent's performance across a wide range of possible environments, which is a more comprehensive and unbiased way of evaluating universal intelligence (Hernández-Orallo & Dowe, 2010). However, defining such a universal distribution in practice is a challenging task. It's not clear how one would go about identifying all possible environments, and the distribution might be also biased towards certain types of environments depending on how it's defined. This indicates that also the measurement of universal intelligence in practice is likely to be an approximation due to the inherent complexities and challenges involved.

3.3 Implications of universal intelligence

This chapter discusses the implications of universal intelligence on the research and understanding of General Intelligence concepts, drawing upon insights from previous sections.

3.3.1 The Singularity

The concept of the 'singularity', as discussed by philosopher David Chalmers, outlines a possible future event where machines exceed human intelligence, leading to rapid technological progress and significant societal changes (Chalmers, 2010). This would also lead to the point in evolution where a species is able to build a system as intelligent as itself (Sutton & Barto, 2018). This idea was further explored by Hutter in his paper 'Can Intelligence Explode?', where he applied this philosophical question to his definition of a universal intelligence (Hutter, 2012a). Hutter proposed that, according to his definition of universal intelligence, it closely correlates with computational limits and the laws of physics, suggesting that humanity could only approach, but never fully reach, the day of the singularity (Hutter, 2012a). This intersects significantly with the theses on computational approximation of intelligence in the AIXI model (see section 3.1.2) and a potential approximation problem from testing or measuring the intelligence of a universal intelligence (see section 3.2.3).

3.3.2 Universal intelligence safety and ethical concerns

The potential realization of a universal intelligence system raises significant safety and ethical considerations. As Hutter indicates, this could precipitate an 'intelligence explosion,' a concept originally proposed by I. J. Good in 1965 (Hutter, 2012a; Good, 1966). In this scenario, a universal intelligence might develop progressively advanced systems, potentially surpassing human intelligence (Hutter, 2012a). The considerable power inherent in universal intelligence magnifies the power of these issues, highlighting the risk of negative outcomes from misuse or inadequate control. This underscores the need for careful preparation and proactive consideration of safety measures and ethical guidelines as researchers should keep a 'safety mindset', a stance echoed by both the European Union and the IEEE (Parliament, for Parliamentary Research Services, Fox-Skelly, Bird, Jenner, Winfield, Weitkamp, & Larbey, 2020).

José Hernández-Orallo (Hernández-Orallo & Dowe, 2010) has suggested that the concept of universal intelligence could be instrumental in creating more advanced systems, such as more intelligent CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) systems that can more effectively distinguish between a human and artificial intelligence (Hernández-Orallo & Dowe, 2010). Additionally, for example universal intelligence tests (see section 3.2.3) could be used to measure progress towards the 'Singularity' (see section 3.3.1) (Hutter, 2012a; Hernández-Orallo & Dowe, 2010). Eliezer Yudkowsky dramatized these statements by combining the explained thesis from section 3.1.2 and ethical concerns developing a potential AGI, asserting, 'Every 18 months, the minimum IQ to destroy the world drops by one point.' (Yudkowsky, 2012). This statement is grounded in Moore's Law, named after Gordon Moore, which posits that the number of transistors on integrated circuits with minimal component costs regularly doubles (Moore, 1965; Waldrop, 2016). According to an assessment by David House, this leads to a doubling of computing power every 18 months, as the semiconductor or chip industry can produce more powerful processors by reducing the size of transistors (Moore, 1965; Kanellos, 2003; Waldrop, 2016). However, as semiconductor sizes approach their physical boundaries, potentially moving beyond 1nm, the anticipated violation of Moore's Law and the emergence of a computational limit is a topic widely discussed by researchers (Yudkowsky, 2012; Ilatikhameneh, Ameen, Novakovic, Tan, Klimeck, & Rahman, 2016). This limitation is possibly aligned with quantum electrodynamics (QED), which represents the most sophisticated model of physics currently known to us (Hutter, 2009).

4. Conclusion

In conclusion, it can be stated that the concept of universal intelligence, as per the theoretical definition by Marcus Hutter, the testing of universal intelligence according to Shane Legg's definition, and the resulting implications are heavily dependent on the discourse surrounding compression. The practical implementation of universal intelligence is in its current state limited by computational resources and the laws of physics. Nonetheless, the insights gained allow for intriguing theoretical conclusions on how universal intelligence can be defined as a concept of intelligence, the advantages that can be derived from it, the compromises that must be made, and how to potentially deal with increasingly powerful general AI systems.

References

Chalmers, D. J. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies*, *17*(9-10), 9–10.

Good, I. J. (1966). Speculations concerning the first ultraintelligent machine. In Alt, F. L., & Rubinoff, M. (Eds.), *Speculations Concerning the First Ultraintelligent Machine*, Vol. 6 of *Advances in Computers*, pp. 31–88. Elsevier.

Hernández-Orallo, J., & Dowe, D. L. (2010). Measuring universal intelligence: Towards an anytime intelligence test. *Artif. Intell.*, *174*(18), 1508–1539.

Hernández-Orallo, J., & Dowe, D. L. (2010). Measuring universal intelligence: Towards an anytime intelligence test. *Artificial Intelligence*, *174*(18), 1508–1539.

Hutter, M. (2000). A theory of universal artificial intelligence based on algorithmic complexity..

Hutter, M. (2005). *Universal Artificial Intelligence*. Texts in Theoretical Computer Science. An EATCS Series. Springer, Berlin and Heidelberg.

Hutter, M. (2006). On the foundations of universal sequence prediction. In Cai, J.-Y., Cooper, S. B., & Li, A. (Eds.), *Theory and Applications of Models of Computation*, pp. 408–420, Berlin, Heidelberg. Springer Berlin Heidelberg.

Hutter, M. (2007a). On universal prediction and bayesian confirmation. *Theoretical Computer Science*, *384*(1), 33–48. Theory and Applications of Models of Computation.

Hutter, M. (2007b). Universal algorithmic intelligence: A mathematical top→down approach. In Goertzel, B., & Pennachin, C. (Eds.), *Artificial General Intelligence*, Cognitive Technologies, pp. 227–290. Springer, Berlin.

Hutter, M. (2009). A complete theory of everything (will be subjective). *CoRR*, *abs/0912.5434*.

Hutter, M. (2012a). Can intelligence explode?..

Hutter, M. (2012b). One decade of universal artificial intelligence..

Hutter, M. (2020a). 500'000 euro prize for compressing human knowledge. http://prize.hutter1.net. (Accessed on 01/01/2024).

Hutter, M. (2020b). Marcus hutter: Universal artificial intelligence, aixi, and agi — lex fridman podcast #75 - youtube. https://www.youtube.com/watch?v=E1AxVXt2Gv4. (Accessed on 01/01/2024).

Ilatikhameneh, H., Ameen, T., Novakovic, B., Tan, Y., Klimeck, G., & Rahman, R. (2016). Saving moore's law down to 1 nm channels with anisotropic effective mass. *Scientific Reports*, *6*(1).

Kanellos, M. (2003). Moore's law to roll on for another decade - cnet. https://www.cnet.com/tech/tech-industry/moores-law-to-roll-on-for-another-decade/. (Accessed on 01/01/2024).

Katayama, S. (2019). Computable variants of aixi which are more powerful than aixitl..

Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence..

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*, *38*(8).

Parliament, E., for Parliamentary Research Services, D.-G., Fox-Skelly, J., Bird, E., Jenner, N., Winfield, A., Weitkamp, E., & Larbey, R. (2020). *The ethics of artificial intelligence – Issues and initiatives*. European Parliament.

Rathmanner, S., & Hutter, M. (2011). A philosophical treatise of universal induction. *CoRR*, *abs/1105.5721*.

Schmidhuber, J. (2016). Künstliche intelligenz wird alles ändern - youtube. https://www.youtube.com/watch?v=rafhHIQgd2A. (Accessed on 01/01/2024).

Sedgewick, R., & Wayne, K. (2016). Intractability. https://algs4.cs.princeton.edu/66intractability/. (Accessed on 01/01/2024).

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (Second edition). The MIT Press.

Veness, J., Ng, K. S., Hutter, M., Uther, W., & Silver, D. (2010). A monte carlo aixi approximation..

Waldrop, M. M. (2016). The chips are down for moore's law. *Nature News*, *530*(7589), 144.

Yudkowsky, E. (2012). Yudkowsky's law - lesswrong. https://www.lesswrong.com/tag/yudkowskys-law. (Accessed on 01/01/2024).