
"Approximately Universally Intelligent" — How AIXI can Inspire Practical AI Algorithms

Nikolas Kirschstein

Technical University of Munich, Germany
nikolas.kirschstein@tum.de

Abstract

Universal Artificial Intelligence (UAI) is a mathematical framework that produces a provably optimal and maximally general agent called AIXI, serving as a theoretical gold standard for artificial general intelligence. However, AIXI is inherently incomputable and thus cannot be implemented directly in practice. Under the lens of generality, this report discusses the merits and limitations of UAI as well as practicable approximations thereof. Overall, UAI provides strong theoretical optimality guarantees and flexible framework modularity that allows adopters to, in principle, "choose" the level of generality. However, the framework makes potentially generality-narrowing modelling assumptions and the few existing AIXI approximations lack quality guarantees with respect to generality. UAI thus remains mainly a theoretical subject for now, with a myriad of open research avenues to explore before it becomes evident whether UAI is the ultimate path to AGI.

1 Introduction

Artificial General Intelligence (AGI) has been a moonshot goal at least since the advent of computers and the Turing Test [Turing, 1950]. While the initial focus was on symbolic approaches like expert systems, other techniques such as neuronal simulation, genetic algorithms, and cognitive architectures were developed over time. Recently, data-driven machine learning approaches have achieved remarkable results, as evidenced by the success of large language models (LLMs). Anyhow, those approaches are mostly built bottom-up, with only few conceptual foundations.

Universal Artificial Intelligence (UAI; Hutter [2000]) fills this gap by providing a rigorous mathematical framework compatible with machine learning that allows for the top-down creation of a parameterless, fully *generally* intelligent decision maker called AIXI that behaves provably optimally in unknown environments. However, the operation of AIXI is incomputable in the Turing sense. Consequently, practical implementation requires approximating components of the framework. In this report, we discuss different AIXI approximations and how much "loss of generality" they incur, in addition to the overall generality benefits and drawbacks of the UAI framework.

Another notable top-down approach to AGI is PowerPlay, a line of work by Schmidhuber [2011, 2013]. There, an increasingly powerful problem solver is constructed via a search problem over pairs of new tasks and modifications to the existing problem solver. Yet, to our knowledge, there are no practical instantiations of this framework, so that we restrict ourselves to UAI in this report.

As a prerequisite, Section 2 describes how to define the notions of generality and universality. Section 3 explains the formal framework of UAI, which is based on decision theory and Bayesian sequence prediction. Section 4 discusses the theoretical shortcomings of UAI as well as potential fixes. Section 5 investigates several UAI approximations, including Monte Carlo Tree Search, Context Tree Weighting, and Feature Reinforcement Learning. Section 6 concludes the report with a summary and an outlook on future work.

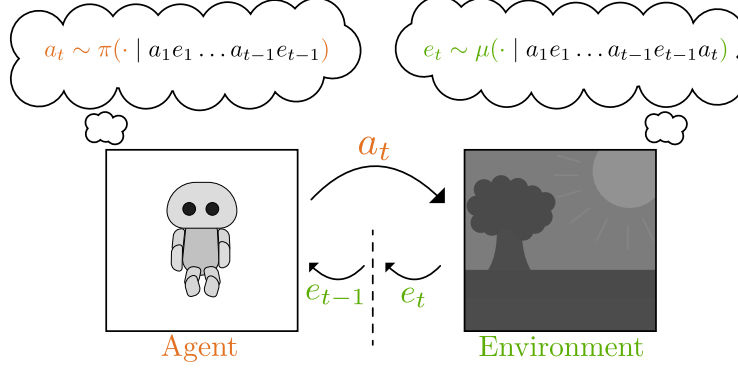


Figure 1: The cybernetic model [based on Hutter et al., 2024, Fig. 6.1].

2 Defining Universal Intelligence

The formal setup for UAI as originally devised by Hutter [2000] is the widely used cybernetic model (cp. Figure 1) that also underlies reinforcement learning (RL) and robotics problems. It is a simple action-reaction paradigm based on the duality of two objects: a policy and an environment. Formally, we need a finite **action space** $\mathcal{A}$ and a finite **percept space** $\mathcal{E} := \mathcal{O} \times \mathcal{R}$ factored into an observation space $\mathcal{O}$ and reward space $\mathcal{R}$.

We collect all possible action-percept sequences in the set $\mathcal{H} := (\mathcal{A} \times \mathcal{E})^*$ of *histories* of the form $h_{\leq m} = a_1 e_1 a_2 e_2 \dots a_t e_t$, where $*$ denotes the Kleene star operation¹. An **agent** or **policy** $\pi : \mathcal{H}^* \rightarrow \Delta \mathcal{A}$ is a probability distribution over actions conditioned on a history. An **environment** $\mu : \mathcal{H} \times \mathcal{A} \rightarrow \Delta \mathcal{E}$ is a probability distribution over percepts conditioned on a history plus action. The agent and the environment interact as follows: start with an empty history $h_0 = ()$ and at each time step $t = 1, 2, \dots$

1. Take an **action** $a_t \in \mathcal{A}$ following the history-dependent (i.e., non-Markov!) **policy** π and
2. Receive a **percept** $e_t = (o_t, r_t) \in \mathcal{E}$ from the **environment** μ .

These alternating dynamics induce a probability measure on histories via the chain rule of probability, analogously to the concept of trajectories in conventional RL, which Hutter et al. formalise as follows.

Definition 1 (Agent-Environment Measure μ^π). For a history $h_{\leq m} = h_{<t} h_{t:m} \in \mathcal{H}$, the joint *agent-environment* measure is given by

$$\mu^\pi(h_{t:m} \mid h_{<t}) := \prod_{k=t}^m \pi(a_k \mid h_{<k}) \cdot \mu(e_k \mid h_{<t}, a_k).$$

The crucial difference to conventional RL, however, is that policy and environment are conditioned on the entire action-observation history, and no Markov and ergodicity assumptions are made. Instead, the framework is kept maximally general, apart from the finiteness assumption on the action and percept spaces and the restriction to a sequential action-reaction procedure. The finiteness assumptions on the underlying spaces can be lifted, resulting in more convoluted notation, but for all practical purposes those spaces are eventually finite on some real computer anyway.

However, the narrow confinement to an agent-environment paradigm may be criticised as a rather simplistic, robotics-motivated RL-style definition, which is taken as a given in the theory of universal intelligence. This can be seen as an inductive bias on a meta-level of the theory of science. However, the cybernetic model has afforded remarkable practical results, and there does not appear to be any work on general intelligence outside of this model, so that we stick to it in this report.

Within the framework of the cybernetic model, most choices of policy would lead to rather random and intuitively "unintelligent" behaviour. To obtain a truly generally intelligent policy, we need to first substantiate these terms.

¹For any set $\mathcal{X}$, its Kleene closure is $\mathcal{X}^* := \bigcup_{i=0}^{\infty} \mathcal{X}^i = \{(x_1, x_2, \dots, x_n) \mid n \in \mathbb{N}; x_1, x_2, \dots, x_n \in \mathcal{X}\}$.

Definition 2 (Intuitive Intelligence). Intelligence is "an agent's ability to achieve goals in a wide range of environments." [Legg and Hutter, 2007]

This definition is rather vague. When accepting the premises of the cybernetic model, at least the terms "agent" and "environment" are well-defined. However, the "wide range" still requires specification. This is done based on the notion of a *universal distribution* over environments, which we will develop later in Definition 7.

Definition 3 (Universal Intelligence). Universal intelligence is "the expected performance of an agent with respect to the universal distribution over the space of all environments." [Legg and Hutter, 2007]

3 Constructing Universal Intelligence

In order to achieve universal intelligence as per Definition 3, we need to construct both an adequate model of the environment as well as the agent itself. Hutter [2000] does this by combining conventional sequential decision theory [Bellman, 1957] – as used in RL – with the theory of universal induction [Solomonoff, 1964] for Bayesian sequence prediction to enable powerful environment modelling. Section 3.1 describes how to model the environment using universal induction, based on which Section 3.2 develops the universally intelligent AIXI agent.

3.1 Environment Model: Universal Induction

In practice, the true environment μ is not known. Hence, Hutter follows the established approach of considering a hypothesis class $\mathcal{V}$ of possible environments ν and using a Bayesian mixture as a proxy for the true, unknown μ .

Definition 4 (Bayes Mixture Environment ξ). Given an environment prior $w : \mathcal{V} \rightarrow \mathbb{R}$ that satisfies $w(\nu) > 0$ for all $\nu \in \mathcal{V}$ and $\sum_{\nu \in \mathcal{V}} w(\nu) \leq 1$, the *Bayes mixture environment* is the weighted sum

$$\xi(h_{\leq m}) := \sum_{\nu \in \mathcal{V}} w(\nu) \cdot \nu(h_{\leq m}). \quad (1)$$

We require strictly positive weights since weighting an environment with a zero prior effectively amounts to using a smaller hypothesis class without that environment. However, since we included that environment in the hypothesis class in the first place, Hutter invokes a philosophical principle to argue that we should not rule it out before seeing evidence that it cannot be the true environment.

Principle 1 (Epicurus's Plurality Principle).

*When multiple models are consistent with the observations, **all of them** should be considered.*

The reason why a Bayes mixture serves as a good proxy for the true environment is that its induced agent-environment measure is guaranteed to converge to the true one in the time limit.

Theorem 1 (Predictive Convergence, informal). *If the true $\mu \in \mathcal{V}$, we have for any policy π that*

$$\xi^\pi(h_{t:m} \mid h_{<t}) \xrightarrow{t \rightarrow \infty} \mu^\pi(h_{t:m} \mid h_{<t}) \quad \text{with probability 1.}$$

Proof. Apply the general sequence prediction result from Hutter [2001a, Theorem 1] to a factored history space and a factorised agent-environment measure. $\square$

To trivially fulfil the precondition, we simply choose $\mathcal{V}$ as the least restrictive but still theoretically achievable hypothesis class: the set of all computable environments, which we note to be countable. Furthermore, while the eventual convergence is independent of the specific choice of prior w , the speed of convergence can be heavily dependent on the prior. To make a reasonable choice of prior, Hutter draws from another basic philosophical principle.

Principle 2 (Occam's Razor).

*When multiple models are consistent with the observations, **only the simplest** should be considered.*

Occam's Razor is arguably the most fundamental principle in all of science. A strong motivation for this simplicity emphasis is that all scientific endeavours can be seen as compressing reality into simple explanations in order to afford predictive power. However, Occam's Razor clashes with

Epicurus's principle and the non-zero prior requirement. One solution is to interpolate between the two principles by assigning to each environment a non-zero weight that is inversely proportional to its complexity. The only remaining issue is then to measure "complexity". While there are many possible choices of complexity measure, a theoretically sound option that Hutter employs is the length of the shortest representation of the environment on a theoretical reference computer.

Definition 5 (Universal Turing Machine [UTM; Turing, 1937]). A *universal Turing machine (UTM)* is a Turing machine $\mathcal{U}$ that *simulates another Turing machine* M given its description $\langle M \rangle$ and an input a , which we concatenate into a program $p := \langle M \rangle a$. In symbols: $\mathcal{U}(p) = \mathcal{U}(\langle M \rangle a) = M(a)$.

Recall from theoretical informatics that UTMs exist but are not unique. Fixing one reference UTM $\mathcal{U}$, we obtain a complexity measure that intuitively quantifies the information content of an object.

Definition 6 (Kolmogorov Complexity [Kolmogorov, 1965]). The *Kolmogorov complexity* of a string x over the alphabet of the UTM $\mathcal{U}$ is the length of the shortest program for $\mathcal{U}$ producing x :

$$K(x) := \min\{|p| : p \text{ is program with } \mathcal{U}(p) = x\}$$

To measure the complexity of an environment ν , we need to encode it as a string over the alphabet of $\mathcal{U}$. Since any arbitrary but fixed bijective encoding can be used, we abuse notation and write $K(\nu)$ for the Kolmogorov complexity of the encoding of ν . Putting everything together, we arrive at the following universal prior over environments, which interpolates between Occam's Razor and Epicurus's Plurality. Environments receive a weight that decreases exponentially with their Kolmogorov complexity.

Definition 7 (The Universal Prior [Solomonoff, 1964]). $w_U(\nu) := 2^{-K(\nu)}$

Definition 8 (The Universal Environment). $\xi_U(h_{\leq m}) := \sum_{\nu \in \mathcal{V}} w_U(\nu) \cdot \nu(h_{\leq m})$

3.2 Agent Model: AIXI

With the environment side of the cybernetic model sorted, we turn to the agent side. Based on the universal environment ξ_U , Hutter constructs a parameterless AI agent that is hence called AI(ξ_U) or simply AIXI. Reminiscent of RL, the optimal policy of this agent is the one that maximises the future expected cumulative discounted reward given the history so far under the model ξ_U^π .

Definition 9 (AIXI Policy). Given history $h_{<t} \in \mathcal{H}$, lifespan $m \in \mathbb{N} \cup \{\infty\}$ and discount function $\gamma : \mathbb{N} \rightarrow (0, \infty)$, usually either $\gamma_k = 1$ for all k or $\gamma_k = 0.99^{k-t}$, an AIXI policy is defined as

$$\pi_{\xi_U}(h_{<t}) := \arg \max_{\pi} \mathbb{E}_{h_{t:m} \sim \xi_U^\pi(\cdot | h_{<t})} \left[\sum_{k=t}^m \gamma_k r_k \right].$$

Theorem 2 (Expectimax Form of AIXI). *Any AIXI policy can be expressed as*

(colour code: *action*, *percept*, *reward*, *UTM*, *program*)

$$\pi_{\xi_U}(h_{<t}) \in \arg \max_{a_t} \sum_{e_t} \cdots \max_{a_m} \sum_{e_m} (r_t + \cdots + r_m) \sum_{p : \mathcal{U}(pa_1 \dots a_m) = e_1 \dots e_m} 2^{-|p|}. \quad (2)$$

Proof. Omitted. See Hutter et al. [2024, Theorem 7.4.2]. □

The expectimax expression (2) affords the interpretation of an alternating game between the AIXI agent and the environment. The agent maximises for the best next action looking ahead into the future up to the lifespan, spanning an expectimax tree (cp. Figure 2). The environment reacts in an "expected" way, corresponding to the summations and weighting by the universal distribution. The red sum expression represents an unnormalised probability that the output of the UTM $\mathcal{U}$ on a random program fed with all future actions matches the sequence of percepts, i.e., the program encoding the environment is consistent with the history so far. In just one line, Equation (2) combines the key ideas of Bayes, Occam, Epicurus, Turing, Bellman, Kolmogorov, and Solomonoff. Overall, the operation of AIXI can be conceptually separated into two parts:

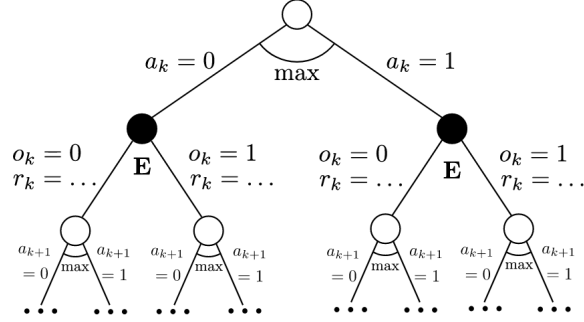


Figure 2: Expectimax tree for binary action/observation spaces [see Hutter et al., 2024, Fig. 12.2].

- **Learning:** Computing the Bayesian mixture ξ_U based on the history so far, through which the agent implicitly updates its belief of what environments in the hypothesis class $\mathcal{V}$ best represent the true environment μ .
- **Planning:** Using ξ_U as a current best estimate of μ to look forward, considering the actions the agent could choose, the percepts that the environment would respond with, the counteractions to each of those percepts, etc., up to the planning horizon, in order to choose the action sequence that would maximise the expected cumulative reward.

This conceptual bisection provides further insight into the construction and will come to bear in the discussion of approximations in Section 5. Before that, we conduct theoretical considerations. Since AIXI is inherently defined by an optimisation process, we can hope for certain optimality properties to hold. Indeed, AIXI is both a Bayes-optimal and a Pareto-optimal agent.

Theorem 3 (Bayes-Optimality of AIXI, informal). *AIXI achieves optimal cumulative reward for the environment model ξ_U .*

Proof. This is guaranteed by Definition 9. □

Theorem 4 (Pareto-Optimality of AIXI, informal). *No other agent achieves at least as much cumulative reward as AIXI in every environment and strictly more reward in at least one environment.*

Proof. Omitted. See Hutter [2000]. □

Due to these optimality results, Hutter et al. [2024] claim that AIXI is the "gold standard for universally optimal agents". So is the UAI framework free from any potential criticism? We address this question in the next section.

4 Scrutinising Universal Intelligence

The narrative of UAI by Hutter et al. is convincing at first glance. However, behind the scenes some theoretical shortcomings of the framework become apparent. In this section, we briefly address those issues and potential fixes for some of them.

4.1 Incomplete Framework

Section 2 already mentioned the assumptions of finite action and percept spaces and sequential action-reaction procedure. The finiteness assumption is not necessary for the theory, but makes it simpler by avoiding involvement in measure-theoretic considerations. The alternating interaction model, however, is heavily restrictive and makes the framework incomplete in several ways.

- **Simultaneity:** The framework does not allow simultaneous or non-alternating interactions between agent and environment by design, even though simultaneity of events is a key feature of real-life dynamics. Yet, since the alternating model imposes significant structure and simplicity, it appears to be an almost unchallenged base assumption in the literature. From the theory of concurrent programs, we know that simultaneity is an intricate mechanic to deal with, and potentially too much of a complexity cost for practical AI approaches.

- **Co-Presence:** The cybernetic model does not take into account the possibility of multiple agents present in the same environment beyond treating them as passive environmental noise. However, since our real-world experience suggests the presence of a large number of intelligent agents, a proper theory of general intelligence should be able to deal with this. Luckily, this has been done extensively in the field of multi-agent RL in the Markovian case and most of the techniques can be adopted in the generic UAI setting as well, as described in the textbook of Hutter et al. [2024, Chapter 10].
- **Reflection:** A cybernetic agent does not consider itself to be part of the environment it interacts with. It only receives percepts and performs actions, but remains static itself and thus lacks self-awareness or a "concept of self". A possible solution to this is so-called reflective AIXI [Leike, 2016] based on the notion of a reflective oracle [Fallenstein et al., 2015], which is limit-computable but not computable and quite involved a construction. An alternative, more practical, and to our knowledge not yet implemented approach would be incorporating an explicit world model into the framework, with representations of itself and the environment (beyond the Solomonoff-induced Bayesian model).

4.2 Reward Design Problem

Just like regular reinforcement learning, UAI suffers from the reward design problem: the reward function, which majorly affects the agent's behaviour, must be specified by the adopter of the framework. While in most game settings the reward design is clear (+1 for winning, -1 for losing), for many even simple tasks, like making a robot learn to walk, it is notoriously hard due to unclear incentive structures and reward hacking by the agent.

In a way, leaving the reward definition up to the algorithm designer is an underspecification of the UAI framework, with the designer thus becoming a teacher to the agent. However, there is a clever fix via an AIXI extension called Knowledge-Seeking Agent [KSA; Orseau et al., 2011, 2013] whose reward function consists of the *knowledge gain* $KL(\xi(\cdot \mid h_{\leq t}) \parallel \xi(\cdot \mid h_{< t}))$ from the last action. This leads to a naturally curious agent behaving like the optimal scientist – always trying to learn as much as possible, which necessitates useful instrumental goals like surviving and memorising. As a result, KSA does not need a teacher like AIXI does and is fully autonomous.

Even with KSA, however, there remains an unsolved alignment problem. It is unknown if (and unlikely that) reward terms exist that instil basic human values like moral behaviour and common sense into an agent. For instance, a KSA deployed in the real world might conclude that its future cumulative information gain is increased when there are no alive humans present in its environment, akin to the well-known paperclip maximiser thought experiment [Bostrom, 2003].

Hence, according to Hutter et al., the ultimate solution would be to start off with a default reward function like knowledge gain and then successively adapt and overwrite it on-the-fly by human-in-the-loop feedback. While this is practical and probably the only viable approach for real-world instantiations of UAI, it is hardly modellable on the theory side.

4.3 Simplicity of Environments

Incorporating the notion of simplicity into the environment prior is a major ingredient to the UAI framework. Schmidhuber [2002] criticises that simplicity does not equate to short program length. Intuitively, simple environments should be easy to compute, but there are many instances where the universal prior w_U places high weight on environments that are computationally complex but have a short generating code.

Instead, Schmidhuber proposes to employ the *speed prior*, which assigns prior weights by computational complexity. In particular, the prior weight of a history is inversely proportional to the required asymptotic runtime of the fastest algorithm producing it. Next to the theoretical appeal of this prior choice, a significant advantage of the speed prior over the universal prior is being recursively approximable to arbitrary precision. Approximability considerations are treated in-depth in the following Section 5.

5 Approximating Universal Intelligence

Next to the theoretical shortcomings, UAI also has severe practical issues that we discuss in this section. Looking at Equation (2), each of the two components of AIXI suffers from a fundamental problem that makes it impractical.

- ① **Incomputability of universal environment (Learning):** The mixture environment ξ_U (Definition 8) uses Kolmogorov complexity, which is known to be uncomputable, and sums over the class $\mathcal{V}$ of all computable environments, which is infinite.
- ② **Intractability of expectimax search (Planning):** The maximisations and summations in the expectimax expression (2) require a search through all possible actions and percepts for all future timesteps, incurring an exponential runtime cost of $\Theta(|\mathcal{A} \times \mathcal{E}|^{m-t+1})$, which is intractable even for moderate values of horizon m .

To make UAI practical, i.e. computable and efficiently so, approximations of AIXI are inevitable. However, the resulting methods will exhibit both theoretical issues and practical issues since we lose theoretical guarantees. So why should we then even deal with all the complexity of UAI and not just go with bottom-up approaches like LLMs that only have theoretical issues but are highly practicable? The answer is that we start with something provably correct (AIXI) and then make it practicable instead of starting with something practicable (LLMs) but inherently wrong in the first place. As a result, the goal of all AIXI approximations must be to retain as much of the theoretical optimality of AIXI as possible while achieving a feasible procedure.

Section 5.1 surveys the prominent approximation techniques for the learning part ①, while Section 5.2 does the same for the planning issue ②. Section 5.3 then introduces the framework of Feature Reinforcement Learning that approaches both issues holistically. As the details of all discussed approximations are drastically technical, we restrict ourselves to high-level descriptions.

5.1 Learning Approximation

Time and Space Bounding. A straightforward approximation of universal induction was proposed together with AIXI already by Hutter [2000, 2001b]. It consists of restricting the space of environment programs p to sum over in Equation (2) to only those of length at most l and for which the UTM halts within t time steps, resulting in a computable approximation called AIXI^{tl} . Hutter shows that it performs at least as well as any other time and space constrained agent. The advantage of AIXI^{tl} is that the designer can directly control the tradeoff between approximation quality and computational complexity via t and l . However, AIXI^{tl} turns out to be practically feasible only for extremely small choices of t and l , in a regime where not even remnants of optimal behaviour can be expected.

Data Compressor Ensemble. Since the incomputability of AIXI arises from the use of Kolmogorov complexity to quantify the simplicity of candidate environments, another straightforward approximation idea consists of replacing it with a computable complexity measure. Instead of employing the idealised Kolmogorov complexity that relies on theoretically optimal compression, one can quantify complexity based on standard data compressors that are efficient. Consequently, the UTM dependence is eliminated from the framework; it becomes computable. However, the quality of the alternative compressor directly affects how closely the prior still resembles Occam’s razor. To retain a powerful environment model, Veness et al. [2012] devise several ensembling techniques such as model averaging, switching, and convex mixing. While practical, the alternative compression approach allows no quantification of approximation quality and empirical performance evaluations only exist in the distinct context of information-based clustering where data compressors replace Kolmogorov complexity in the similarity measure [Cilibrasi and Vitanyi, 2005].

Context Tree Weighting. A different angle of attack on ① is to reduce the class of considered environment models to k -Markov environments, i.e., those that only depend on at most k recent history items. This results in a finite and thus computable set of candidate environments. However, their number is doubly-exponential in k , namely $\Theta(|\mathcal{E}|^{|\mathcal{A} \times \mathcal{E}|^k})$, and hence prohibitively large. To achieve feasible Bayesian inference over environments, the Context Tree Weighting [CTW; Willems et al., 1993, 1995] algorithm can be employed. Sunehag and Hutter [2013] describe how to adapt it for the AIXI learning setting, where CTW allows for updating the mixture weights in $\Theta(k)$ time, a remarkable speedup over naïve inference. The only drawback of this approach, called AIXI-CTW, is a significant loss of generality due to the k -Markov assumption. For instance, a hypothetical environment with a transformative event in the distant past that the agent must remember to be able

to survive would be outside the k -Markov assumption. To mitigate this, Yang-Zhao et al. [2022] introduce Φ -AIXI-CTW that uses logical state abstractions of the history instead of truncating it to k most recent events. However, this approach requires the manual generation of predicates over the history space, with no guarantee that inherently relevant predicates are included.

5.2 Planning Approximation

Monte Carlo Tree Search. Computing all maximums and sums as per Equation (2) corresponds to a brute-force search in the exponentially growing expectimax tree. If exact planning is desired, this is the best we can do. However, if we allow for a stochastic approach to planning, the computational cost can be significantly reduced by repeatedly sampling paths from the expectimax tree instead of traversing all of them. This idea is called Monte Carlo Tree Search [Coulom, 2007] and was adapted by Veness et al. [2010, 2011, 2012] for AIXI planning. They improve efficiency by developing a generalisation of the Upper Confidence Tree [UCT; Kocsis and Szepesvári, 2006] algorithm called ρ -UCT, which was a crucial component in the success of DeepMind’s AlphaGo [Silver et al., 2016]. A considerable advantage of Monte Carlo planning is that the designer can control approximation quality via the number of samples to draw. In the limit of infinite samples, Monte Carlo search becomes exact search. However, beyond toy problems like Go, a large number of samples is required to successfully navigate complex environments and compute becomes a bottleneck once more.

Quantum AIXI. With the increasing progress in quantum computing, Perrier [2025] develops a quantum formulation of AIXI called QAIXI. While the incomputability of Kolmogorov complexity carries over to quantum computers, the planning part holds promise for a significant quantum speedup. The conditions under which such a quantum advantage would exist are not known yet and are a subject of ongoing research. However, Perrier already determines that the necessary quantum error correction for implementation will require millions of physical qubits, suggesting no practical relevance of QAIXI until a point in the future where such an amount of qubits can be technically materialised.

5.3 Feature Reinforcement Learning

Instead of aiming at separate approximations for learning and planning in AIXI, it is reasonable to also consider an integrated approach. One increasingly prominent such approach is Feature Reinforcement Learning [FRL; Hutter, 2009], which aims to compress the full history space $\mathcal{H}$ into an abstract state space $\mathcal{S}$ via a *feature map* $\Phi : \mathcal{H} \rightarrow \mathcal{S}$ and solve the thus induced MDP with standard RL methods. The idea is that, if Φ is carefully constructed, the agent-environment interaction with respect to the features extracted by Φ is approximately Markovian. Note that we do not assume that the true dynamics are Markovian but only that they have a Markov *encoding*. Since efficient methods for standard MDPs exist, the entire problem reduces to automatically inferring a suitable Φ resembling such Markov encoding. The theoretical ideal would be the coarsest possible feature map that still supports optimal decision-making, called *extreme state aggregation* [ESA; Hutter, 2014]. Unfortunately, ESA turns out to be incomputable, analogously to Kolmogorov complexity in vanilla AIXI. Hence, practical approximations are once again necessary, and the most established approach so far is using Context Tree Maximisation [CTM; Nguyen et al., 2012] for history compression to make the feature map tractable. Still, FRL is in its infancy and a subject of ongoing research.

6 Conclusion

This report discussed UAI, the furthest-developed mathematical framework capturing maximally general artificial intelligence. It combines Bayesian inference as a notion of from-experience learning with sequential decision theory as a notion of world interaction, weaving in the philosophical principles of Occam and Epicurus. The result is AIXI, a provably optimal *generally* intelligent cybernetic agent. UAI allows a top-down high-level view on AGI that prevents algorithm designers from micro-organising details and narrowing down their systems’ generality too early. Many components of the framework can be flexibly adapted from being generic to use case-specific. However, there are two significant limitations of UAI: First, the abstract framework is impractical and requires approximations, on whose generality loss hardly any quantifications or guarantees exist. Second, UAI inherits all generality-related weaknesses of the cybernetic model it piggybacks onto, above all the lack of simultaneity, co-presence, and reflection. Addressing both of these limitations should be the subject of future work on this field. Examples like AlphaGo show that UAI and AIXI are capable of inspiring the design of practical AI algorithms that are certainly more principled and potentially more general than bottom-up approaches like LLMs, rendering UAI a promising path to AGI.

References

- Richard Bellman. *Dynamic Programming*. Dover Publications, 1957. ISBN 978-0-486-42809-3.
- Nick Bostrom. Ethical Issues in Advanced Artificial Intelligence. In *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*, volume 2, pages 12–17. 2003. ISBN 1-894613-73-2.
- R. Cilibrasi and P.M.B. Vitanyi. Clustering by Compression. *IEEE Transactions on Information Theory*, 51(4):1523–1545, April 2005. ISSN 0018-9448. doi: 10.1109/TIT.2005.844059.
- Rémi Coulom. Efficient Selectivity and Backup Operators in Monte-Carlo Tree Search. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, H. Jaap Van Den Herik, Paolo Ciancarini, and H. H. L. M. Donkers, editors, *Computers and Games*, volume 4630, pages 72–83. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. ISBN 978-3-540-75537-1 978-3-540-75538-8. doi: 10.1007/978-3-540-75538-8_7.
- Benja Fallenstein, Jessica Taylor, and Paul F. Christiano. Reflective Oracles: A Foundation for Game Theory in Artificial Intelligence. In Wiebe Van Der Hoek, Wesley H. Holliday, and Wenfang Wang, editors, *Logic, Rationality, and Interaction*, volume 9394, pages 411–415. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015. ISBN 978-3-662-48560-6 978-3-662-48561-3. doi: 10.1007/978-3-662-48561-3_34.
- Marcus Hutter. A Theory of Universal Artificial Intelligence based on Algorithmic Complexity, 2000.
- Marcus Hutter. Convergence and Error Bounds for Universal Prediction of Nonbinary Sequences. In G. Goos, J. Hartmanis, J. Van Leeuwen, Luc De Raedt, and Peter Flach, editors, *Machine Learning: ECML 2001*, volume 2167, pages 239–250. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001a. ISBN 978-3-540-42536-6 978-3-540-44795-5. doi: 10.1007/3-540-44795-4_21.
- Marcus Hutter. Towards a Universal Theory of Artificial Intelligence Based on Algorithmic Probability and Sequential Decisions. In G. Goos, J. Hartmanis, J. Van Leeuwen, Luc De Raedt, and Peter Flach, editors, *Machine Learning: ECML 2001*, volume 2167, pages 226–238. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001b. ISBN 978-3-540-42536-6 978-3-540-44795-5. doi: 10.1007/3-540-44795-4_20.
- Marcus Hutter. Feature Reinforcement Learning: Part I. Unstructured MDPs. *Journal of Artificial General Intelligence*, 1(1), January 2009. ISSN 1946-0163. doi: 10.2478/v10229-011-0002-8.
- Marcus Hutter. Extreme State Aggregation beyond MDPs. In Peter Auer, Alexander Clark, Thomas Zeugmann, and Sandra Zilles, editors, *Algorithmic Learning Theory*, volume 8776, pages 185–199. Springer International Publishing, Cham, 2014. ISBN 978-3-319-11661-7 978-3-319-11662-4. doi: 10.1007/978-3-319-11662-4_14.
- Marcus Hutter, David Quarel, and Elliot Catt. *An Introduction to Universal Artificial Intelligence*. Chapman and Hall/CRC, Boca Raton, 1 edition, April 2024. ISBN 978-1-003-46029-9. doi: 10.1201/9781003460299.
- Levente Kocsis and Csaba Szepesvári. Bandit Based Monte-Carlo Planning. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Johannes Fürnkranz, Tobias Scheffer, and Myra Spiliopoulou, editors, *Machine Learning: ECML 2006*, volume 4212, pages 282–293. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. ISBN 978-3-540-45375-8 978-3-540-46056-5. doi: 10.1007/11871842_29.
- Andrei N Kolmogorov. Three approaches to the quantitative definition of information. *Problems of information transmission*, 1(1):1–7, 1965.
- Shane Legg and Marcus Hutter. Universal Intelligence: A Definition of Machine Intelligence, December 2007.

- Jan Leike. Nonparametric General Reinforcement Learning, 2016.
- Phuong Nguyen, Peter Sunehag, and Marcus Hutter. Context tree maximizing reinforcement learning. In *AAAI-12 / IAAI-12 - Proceedings of the 26th AAAI Conference on Artificial Intelligence and the 24th Innovative Applications of Artificial Intelligence Conference*, Proceedings of the National Conference on Artificial Intelligence, pages 1075–1082, 2012. ISBN 978-1-57735-568-7.
- Laurent Orseau. Universal Knowledge-Seeking Agents. In Jyrki Kivinen, Csaba Szepesvári, Esko Ukkonen, and Thomas Zeugmann, editors, *Algorithmic Learning Theory*, volume 6925, pages 353–367. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-24411-7 978-3-642-24412-4. doi: 10.1007/978-3-642-24412-4_28.
- Laurent Orseau, Tor Lattimore, and Marcus Hutter. Universal Knowledge-Seeking Agents for Stochastic Environments. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Sanjay Jain, Rémi Munos, Frank Stephan, and Thomas Zeugmann, editors, *Algorithmic Learning Theory*, volume 8139, pages 158–172. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-40934-9 978-3-642-40935-6. doi: 10.1007/978-3-642-40935-6_12.
- Elija Perrier. Quantum AIXI: Universal Intelligence via Quantum Information, 2025.
- Jürgen Schmidhuber. The speed prior: A new simplicity measure yielding near-optimal computable predictions. In Jyrki Kivinen and Robert H. Sloan, editors, *Computational Learning Theory*, pages 216–228, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg. ISBN 978-3-540-45435-9. doi: 10.1007/3-540-45435-7_15.
- Jürgen Schmidhuber. POWERPLAY: Training an Increasingly General Problem Solver by Continually Searching for the Simplest Still Unsolvable Problem, 2011.
- Jürgen Schmidhuber. PowerPlay: Training an Increasingly General Problem Solver by Continually Searching for the Simplest Still Unsolvable Problem. *Frontiers in Psychology*, 4, 2013. ISSN 1664-1078. doi: 10.3389/fpsyg.2013.00313.
- David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, January 2016. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature16961.
- R.J. Solomonoff. A formal theory of inductive inference. Part I. *Information and Control*, 7(1):1–22, March 1964. ISSN 00199958. doi: 10.1016/S0019-9958(64)90223-2.
- Peter Sunehag and Marcus Hutter. Context Tree Weighting. 2013.
- A. M. Turing. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1):230–265, 1937. ISSN 00246115. doi: 10.1112/plms/s2-42.1.230.
- A. M. Turing. Computing Machinery and Intelligence. *Mind*, LIX(236):433–460, October 1950. ISSN 1460-2113, 0026-4423. doi: 10.1093/mind/LIX.236.433.
- J. Veness, K.S. Ng, M. Hutter, W. Uther, and D. Silver. A Monte-Carlo AIXI Approximation. *Journal of Artificial Intelligence Research*, 40:95–142, January 2011. ISSN 1076-9757. doi: 10.1613/jair.3125.
- Joel Veness, Kee Siong Ng, Marcus Hutter, and David Silver. Reinforcement Learning via AIXI Approximation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 24(1):605–611, July 2010. ISSN 2374-3468, 2159-5399. doi: 10.1609/aaai.v24i1.7667.

- Joel Veness, Peter Sunehag, and Marcus Hutter. On Ensemble Techniques for AIXI Approximation. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Joscha Bach, Ben Goertzel, and Matthew Iklé, editors, *Artificial General Intelligence*, volume 7716, pages 341–351. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. ISBN 978-3-642-35505-9 978-3-642-35506-6. doi: 10.1007/978-3-642-35506-6_35.
- F.M.J. Willems, Y.M. Shtarkov, and T.J. Tjalkens. Context Tree Weighting : A Sequential Universal Source Coding Procedure for Fsmx Sources. In *Proceedings. IEEE International Symposium on Information Theory*, pages 59–59, San Antonio, TX, 1993. IEEE. ISBN 978-0-7803-0878-7. doi: 10.1109/ISIT.1993.748374.
- F.M.J. Willems, Y.M. Shtarkov, and T.J. Tjalkens. The context-tree weighting method: Basic properties. *IEEE Transactions on Information Theory*, 41(3):653–664, May 1995. ISSN 00189448. doi: 10.1109/18.382012.
- Samuel Yang-Zhao, Tianyu Wang, and Kee Siong Ng. A direct approximation of AIXI using logical state abstractions. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 36640–36653. Curran Associates, Inc., 2022.