

The Ecosystem Path to General AI

Claes Strannegård*, Niklas Engsner, Pietro Ferrari, Hans Glimmerfors, Marcus Hilding Södergren, Tobias Karlsson, Birger Kleve, and Victor Skoglund

Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
Sweden

Abstract. We start by discussing the link between ecosystem simulators and general AI. Then we present the open-source ecosystem simulator Ecotwin, which is based on the game engine Unity and operates on ecosystems containing inanimate objects like mountains and lakes, as well as organisms such as animals and plants. Animal cognition is modeled by integrating three separate networks: (i) a *reflex network* for hard-wired reflexes; (ii) a *happiness network* that maps sensory data such as oxygen, water, energy, and smells, to a scalar happiness value; and (iii) a *policy network* for selecting actions. The policy network is trained with reinforcement learning (RL), where the reward signal is defined as the happiness difference from one time step to the next. All organisms are capable of either sexual or asexual reproduction, and they die if they run out of critical resources. We report results from three studies with Ecotwin, in which natural phenomena emerge in the models without being hardwired. First, we study a terrestrial ecosystem with wolves, deer, and grass, in which a Lotka-Volterra style population dynamics emerges. Second, we study a marine ecosystem with phytoplankton, copepods, and krill, in which a diel vertical migration behavior emerges. Third, we study an ecosystem involving lethal dangers, in which certain agents that combine RL with reflexes outperform pure RL agents.

Keywords: ecosystem · neural networks · happiness · reflexes · reinforcement learning.

Animal cognition

All organisms in nature are subject to natural selection and, where applicable, also sexual selection [1]. These forces operate on physical as well as cognitive properties of the organisms. In particular, they operate on intelligence. One factor that contributes to the selection pressure on intelligence is other organisms. In fact, organisms coevolve with each other and are part of each other’s environment. Intelligence is an advantage in the “arms race” between predators and prey, in the competition for scarce resources, in collaborations with mutual benefits, and in sexual selection processes. Another factor that contributes to

* Corresponding author: claes.strannegard@chalmers.se

the selection pressure on intelligence is the terrain. In fact, animals must continuously handle challenges imposed by the terrain and adapt their behavior to local conditions when migrating, foraging, approaching prey, escaping predators, mating, and parenting.

To survive and reproduce, animals need to deal with a continuous stream of challenges in their lives. They must find food, avoid predators, navigate, and mate. Solving these challenges requires efficient information processing, in particular perception, decision-making, and action. Nervous systems are present in almost all taxa of the animal kingdom and play a key role in animal information processing. They are typically far from monolithic. For example, in humans, the complex nervous system that controls digestion is essentially separated from the brain and may operate even in a brain-dead person. The human brain itself is highly modular, with its anatomically distinct lobes and regions, such as the *Brain stem* which controls reflexes —positive, e.g., the knee reflex, and negative, e.g., the diving reflex; the *Prefrontal cortex*, which maps sensory signals to actions; and the *Insula*, which combines internal signals, such as blood sugar level and external signals, such as smells, into signals linked to happiness and reward [2].

Reflexes seem to be critical to many animals [3]. For example, a newborn lamb might not have the time to find its way to its mother’s udder through random trial and error. Instead, its life might be saved by an instinct that draws it to the smell of milk and a positive suck reflex, which causes it to drink. The lamb might also benefit critically from negative reflexes that prevent it from eating lethal objects, inhaling liquid, or jumping from high cliffs. From an evolutionary perspective, one of the great advantages of nervous systems is that they enable learning, i.e., physical modification, and thereby efficient adaptation to the dangers and resources of the local environment. A prominent example of learning is reinforcement learning (RL), which is used across the animal kingdom [4,5]. Many animals combine untrainable reflex circuits with circuits that are trainable with RL. For instance, humans have hundreds of hardwired reflexes, e.g. the knee reflex, but we can also learn via RL. This enables us to combine the benefits of reflexes with the benefits of RL.

Models of animal cognition

There is a great number of computational models of animal cognition. In the “reflex tradition”, animals are modeled as reflex agents, without the ability to learn. Such models might be adequate for animals with hardwired nervous systems that remain essentially unchanged during their lifetime. An example could be the famous nematode *C. elegans* [6]. Many biologically inspired algorithms, such as cellular automata, swarm algorithms, and ant algorithms, belong to this tradition and so does the subfield that studies populations of reflex agents that develop according to evolutionary algorithms.

In the “RL tradition”, animals are modeled as RL agents. This field includes physical animal robots that get rewarded for fast crawling, running, swimming, or flying. It also includes homeostatic agents that get rewarded for keeping a

set of homeostatic variables —like energy, water, and oxygen— close to their target values, or sweet spots [7,8]. RL algorithms have been shown to have a great potential for AI. For instance, they outperform humans at several video games and strategic board games [9]. There are also many agent-based ecosystem simulators that are based on RL [10,11,12].

In analytic approaches to ecosystem modeling, organisms are typically modeled with numbers representing population size or biomass and the interaction dynamics is modeled with systems of differential equations. There is no model of individual animals, no model of animal cognition, and no model of the terrain. Examples include the well-known Lotka-Volterra predator-prey dynamics [13]. The Ecopath (with Ecosim and Ecospace) simulator for marine ecosystems [14] divides maps of ecosystem into geographical cells, where each cell contains populations, for example, given as tonnes of phytoplankton, zooplankton, planktivores, and piscivores.

AI via models of animal cognition

Nervous systems provided the original inspiration for the neural network model and its applications to supervised learning [15] and RL [5]. Some researchers try to copy animal brains in wetware or reproduce their connectome in software [6], others build computational models of the brain, sometimes called cognitive architectures [16]. Rather than aiming directly for a model of the brain, one may aim for a model of the process that led to the development of the brain. Since natural (general) intelligence emerged as the result of an evolutionary process that took place in an ecosystem, a natural strategy for creating (general) AI is to construct an ecosystem simulator that exploits the natural selection pressure on intelligence. This strategy is in line with Wilson’s “animat path to AI” [17] and it enables a gradual approach to AI that starts with relatively simple ecosystems.

Section 1 of this paper presents the theoretical framework of the ecosystem simulator Ecotwin, which combines RL and reflexes. Section 2 presents simulation results, in which well-known patterns from population dynamics and ethology emerge. Section 3, finally, draws some conclusions.

1 Ecosystem simulator

In this section, we present the theoretical framework of the ecosystem simulator Ecotwin. Ecotwin uses the game engine Unity with ML-Agents [10], which provides a graphical user interface, a physics engine, and several RL algorithms. More details about Ecotwin, including its open source code, can be found at www.ecotwin.se.

Ecosystems

We model *space* with a set $\mathcal{S} \subseteq \mathbb{R}^3$ and *time* with the numbers $0, 1, 2, \dots$. The building blocks of our ecosystems are called *objects*. For instance, we could in-

roduce cat objects, dandelion objects, and rock objects. Each object is assigned a set of object properties.

Definition 1 (Object properties). *The object properties are:*

- Physical properties (*measured in their respective SI units*) such as *Temperature, Mass, Pressure, Electric current, and Luminous intensity*. Moreover, each object has a *conformation*, which is a subset of $\mathcal{S}$.
- Chemical properties (*measured in the SI unit molarity*) are concentrations of chemical substances such as *Oxygen, Nitrogen, Water, Carbon dioxide, Glucose, Salt, Sulfur, Fat, Protein, Estrogen, Testosterone, and Oxytocin*.
- Biological properties such as *Age (seconds), Sex (none, female, or male), and Fertility (a real number in $[0, 1]$)*.

Note that the conformation of an object describes its shape and position in $\mathcal{S}$. The physical and chemical properties of objects may be defined as aggregations of the properties of the points inside their conformations, e.g., the average temperature of a rock object. It is sometimes convenient to introduce additional properties. For example, to talk about cat objects, one may want to introduce properties such as Eye color, Paw size, Blood pressure, and Blood sugar.

Definition 2 (Inanimate objects). *An inanimate object in $\mathcal{S}$ consists of*

- *A type: for example Rock, Water, Road, Building*
- *A set of physical and chemical properties*

Definition 3 (Organisms). *An organism in $\mathcal{S}$ consists of*

- *A type: for example a particular species of bacteria, fungus, plant, or animal*
- *A genome consisting of a string over a finite alphabet*
- *A set of physical, chemical, and biological properties*
- *A nervous system consisting of a set of sensors S , where each sensor has a location in the conformation and a sensitivity to some physical or chemical property; a set of actions A , and three disjoint networks: a reflex network with input nodes S and output nodes A ; a policy network with input nodes S and output nodes A ; and a happiness network with input nodes S and a single output node for representing a scalar happiness value.*
- *A set of hyperparameters. These may include hyperparameters for training the policy network and for regulating reproduction and death. An example of a hyperparameter is `age_max`, which controls the maximum lifetime.*
- *A set of update rules:*
 - *Update rules for physical properties, e.g. how locomotion actions influence the conformation.*
 - *Update rules for chemical properties, e.g. how ingestion and locomotion actions influence the Glucose property.*
 - *Update rules for biological properties specifying (i) the reproduction process, which may be sexual or asexual and involve mutation and crossover; and (ii) the physical and chemical properties that must be met for the organism to stay alive. For example, if Temperature is not inside a certain interval, or if `Age` > `age_max`, then the organism will die and thus become an inanimate object.*

The genome has two roles: (1) It encodes the organism at the start of its “life”, in particular, it could encode its initial conformation (body shape), nervous system, and hyperparameters; (2) It is the input to the reproduction process, in which one or two genomes give rise to a new genome.

This definition of organism encompasses pure reflex agents (with an empty policy network), pure RL-agents (with an empty reflex network), and agents that combine reflexes with RL. By leaving the nervous system empty, we can also model agents lacking nervous systems altogether, such as bacteria and plants. It is often convenient to represent the reflex and happiness networks as hand-coded functions. If desired, these functions can easily be converted into neural networks, by using supervised learning. This step might be useful when applying mutations that add noise to connection weights.

Definition 4 (Ecosystem). *An ecosystem consists of*

- A space $\mathcal{S} \subseteq \mathbb{R}^3$
- A set of inanimate objects in $\mathcal{S}$
- A set of organisms in $\mathcal{S}$.

Several examples of ecosystems will be given in Section 2. The ecosystem is updated at each tick, using the physics mechanisms of Unity and the update rules of the organisms.

Decision-making

This is how the nervous system is used for making decisions at time t :

1. For each sensor, read off its physical or chemical property at its current position. This produces a vector $\mathbf{x}$ that encodes the sensory input at t .
2. Give the input $\mathbf{x}$ to the Reflex network, which then outputs some vector $\mathbf{y}$, with component values -1, 0, or 1. The intended meanings of these values are, respectively, to block, accept, and force the corresponding action.
3. Also give the input $\mathbf{x}$ to the Policy network, which then selects an action a . Represent this action as a one-hot vector $\mathbf{z}$, where 1 is in the position of a .
4. Finally, compute the vector $H(\mathbf{y} + \mathbf{z})$. Here, $H((v_1, \dots, v_n))$ is defined as $(h(v_1), \dots, h(v_n))$, where $h(v_i)$ is 1 if $v_i > 0$ and 0 otherwise. This produces a multi-hot vector that constitutes the decision at t . Thus, the decision might be no action, one action, or many actions.

Now that we have seen how nervous systems are used for decision-making, let us consider a concrete example of a nervous system of a lamb model. The sensors, actions, and reflex network are shown in Fig. 1. The happiness network maps Glucose and Oxygen to a happiness value, so that ingestion and breathing are encouraged. An additional input could be Sheep_smell, so that the animal is encouraged to approach sheep: a hardwired social instinct. The policy network is trained using RL and updated at each time step. Thus, the animal might learn to breathe when Oxygen is low and move toward the strongest smell of milk when Glucose is low.

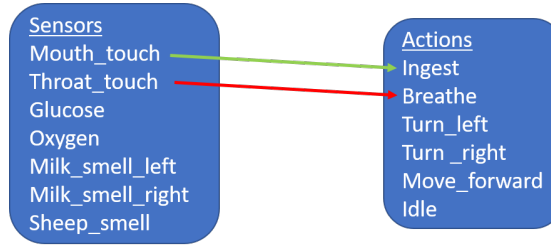


Fig. 1: A Reflex network. This network has sensors for touch (pressure) at two different locations of the conformation. It also has sensors for glucose and oxygen concentration in the blood stream, and for smell at different locations. The actions are for homeostatic regulation and locomotion. Only two connections have non-zero weights: one with weight +1 (green) and one with weight -1 (red). This produces a positive sucking reflex and a negative reflex that is similar to a diving reflex and prevents inhalation of liquid.

Learning

The reflex and happiness networks develop through evolution only, and not through learning. In contrast, the policy network is updated at each time step via RL. The reward signal at time t is defined as $happiness(t) - happiness(t-1)$, where $happiness$ is computed by the happiness network. In the simulations discussed in this paper, we use the standard RL algorithm PPO [18] together with the animals' individual hyperparameters.

2 Results

We will present several results that were obtained using Ecotwin simulations. Videos of the simulations are available at www.ecotwin.se. Ecotwin can run on ecosystems populated with models of real organisms, imaginary organisms, or robots. In this section, we will consider simple models of real organisms. The properties and the mechanisms of our model organisms will be taken from the corresponding real organisms. In order to find a reasonable starting point, we initialize the policy networks randomly and then pre-train them with RL in Ecotwin, before turning on reproduction and starting the simulations.

Emergence of predator-prey dynamics

Previous work in predator-prey dynamics has shown that a two-species predator-prey system, with agents trained through RL, exhibited Lotka-Volterra cycles under certain choices of parameters [12]. We wanted to explore whether a three-species system would also exhibit such cycles.

Our study concerns a three species predator-prey ecosystem with grass, deer, and wolves, as illustrated in Fig. 2 (Left). Deer and wolves have vision, which is

modeled via Unity’s ray casts, and gives the direction and distance to the closest visible objects of each type, within a certain radius [10]). Moreover, the deer can smell grass and wolves, while the wolves can smell deer. At each time step, each animal decides whether it should stand still, walk, or run forward. It also decides whether it should rotate left, right, or not at all. The happiness of the deer is determined by their Energy and Wolf smell sensors, whereas happiness of the wolves is determined by their Energy and Deer smell sensors. The deer obtains energy by eating grass and wolves by eating deer. The consumption of energy depends on the velocity of the animal. A simplified reproduction mechanism was used, where each animal had a probability of giving birth to a new agent, which was then spawned at a random unoccupied location in the ecosystem.

We ran two simulations with Ecotwin, starting with the same ecosystem. The result is shown in Fig. 2 (Right). As expected, given the dependence on randomness, the two simulations are different. In both simulations we see Lotka-Volterra cycles, with an increase in grass, followed by an increase in deer, followed by an increase in wolves, and similar decreases. More details about the study can be found in [19].

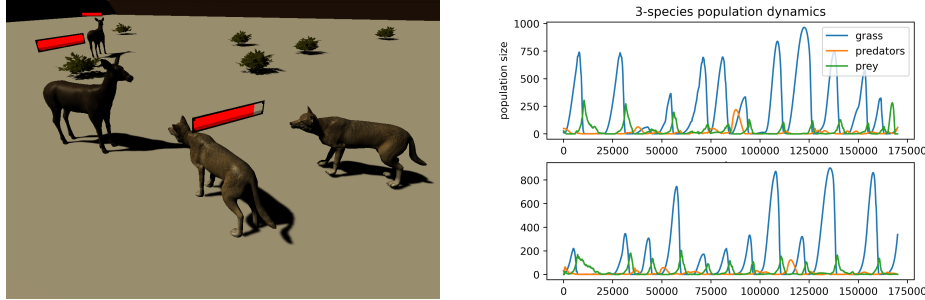


Fig. 2: **Left:** A three-species predator-prey ecosystem, with grass, deer and wolves. The red bars show happiness values. **Right:** Population dynamics from two different simulations.

Emergence of diel vertical migration

In this study, we consider a marine ecosystem with krill, copepods and phytoplankton, illustrated in Fig. 3 (Left). A diel light cycle was simulated, with the sun going up and down. We also simulated decreasing light intensity at greater depths. Three behaviors observed in real Copepods [20] were studied: (1) Diel Vertical Migration (DVM): a cyclic behavior in which Copepods migrate to the surface at night (enabling them to graze phytoplankton that are near the surface where the light is) and go down to greater and darker depths in the daytime (where the vision of predators such as krill is less efficient). (2) Quick escape re-

actions that enable Copepods to escape from predators. (3) Chemotaxis causing Copepods to avoid the smell of predators.

The Copepod models have a simple form of vision, by which they can perceive a rough direction and distance to near-by phytoplankton. Moreover, they can perceive light intensity, their own energy level, and whether they are touching food, krill, the environment’s boundaries, or another Copepod. Their happiness value, yielding their reward, depends on energy and light intensity.

The Copepods and krill were pre-trained with RL for 1 million time steps. Then an Ecotwin simulation was run, starting with the ecosystem shown in Fig. 3 (Left). The behavioral patterns (1)-(3) were observed in the simulations. In particular, a clear DVM pattern was observed, as shown in Fig. 3 (Right). More information about the study can be found in [21].

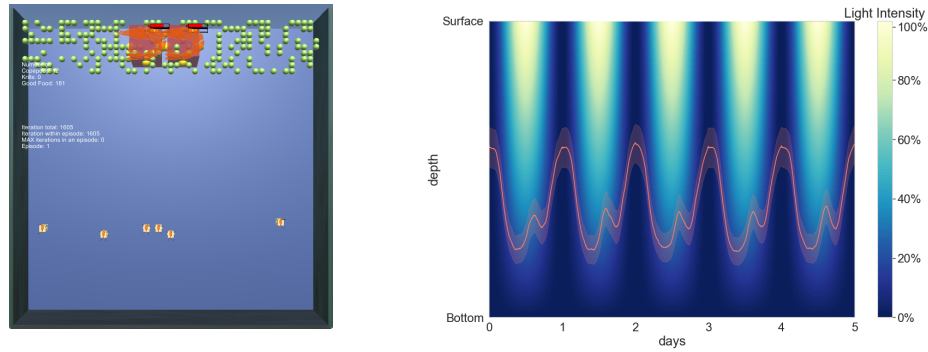


Fig. 3: **Left:** A marine environment with krill (orange), Copepods (yellow), and phytoplankton (green). The light intensity at different depths varies over time. This snapshot was taken at simulated daytime, when the Copepods are relatively far from the surface. **Right:** Mean vertical position of the copepods (red curve) over time. The light intensity is indicated by the background color.

Emergence of critical reflexes

This study concerns the interplay between RL and reflexes. We consider a terrestrial ecosystem with goats and grass, illustrated in Fig. 4 (Left). There are three types of grass in the environment: green, yellow and red. The green grass is good for the goats to eat, the yellow grass is bad, but not deadly, whereas the red grass is deadly. The goats reproduce sexually and have a genome that encodes their policy networks and reflex networks. Thus, they may pass on genomes, including genes that code for reflexes, to their offspring. The goats have four different genes that we call red, yellow, green, and blue, for convenience. These genes correspond to a reflex which prevents a goat from eating red, yellow and green food objects, respectively. The blue gene does not affect a goat’s reflexes.

An Ecotwin simulation was run with these organisms and the result is shown in Fig. 4 (Right). We can see several patterns: (1) The red gene eventually dominates the population. This is expected, as a reflex to avoid lethal food gives them a clear advantage. (2) The domination of the red gene suggests that the combination of reflexes and RL is more effective than RL alone when there are lethal dangers in the environment. Avoiding lethal food cannot be learned during a goat’s life. (3) The yellow gene does not have a clear advantage over the blue gene. This suggests that the combination of reflexes and RL gives no advantage, that cannot be learned, over pure RL-agents when the dangers are not lethal. (4) The goats with the green gene keep dying out and then reappearing because of the mutation in the inheritance mechanisms. (5) A goat’s genome may include multiple genes. This explains why the superior red gene does not dominate the population completely. More details about the study can be found in [22].

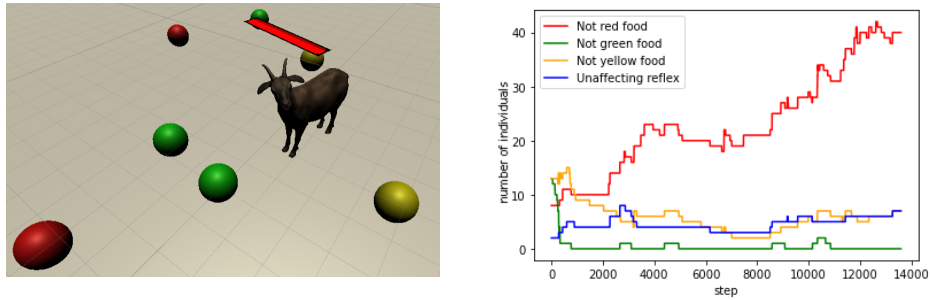


Fig. 4: **Left:** An ecosystem with goats, and three different types of grass objects: green = good, yellow = bad but non-lethal, and red = lethal. **Right:** The number of goats that have the red, yellow, green, and blue genes.

3 Conclusion

We have presented the open source ecosystem simulator Ecotwin. It was run on three ecosystems, on which it reproduced certain population dynamics and behavioral patterns that can be observed in real ecosystems. Agent-based ecosystem simulators can be used for predicting the consequences of human interventions, e.g., via fishing, forestry, or urbanization. They might also be used as “general AI gyms”, in which populations of agents coevolve in a fully automatic process, while taking advantage of the built-in selection pressure on intelligence. This research path toward general AI seems to be worthy of further exploration.

References

1. C. Darwin, “On the Origin of Species by Means of Natural Selection,” 1859. The complete work of Charles Darwin online: darwinonline.org.uk.

2. R. Nieuwenhuys, “The insular cortex: A review,” in *Evolution of the Primate Brain* (M. A. Hofman and D. Falk, eds.), vol. 195 of *Progress in Brain Research*, pp. 123–163, Elsevier, 2012.
3. R. B. Stein and C. Capaday, “The modulation of human reflexes during functional motor tasks,” *Trends in neurosciences*, vol. 11, no. 7, pp. 328–332, 1988.
4. Y. Niv, “Reinforcement learning in the brain,” *Journal of Mathematical Psychology*, vol. 53, no. 3, pp. 139–154, 2009.
5. E. O. Neftci and B. B. Averbeck, “Reinforcement learning in artificial and biological systems,” *Nature Machine Intelligence*, vol. 1, no. 3, pp. 133–143, 2019.
6. C. S. Xu, M. Januszewski, Z. Lu, S.-y. Takemura, K. Hayworth, G. Huang, K. Shinomiya, J. Maitin-Shepard, D. Ackerman, S. Berg, *et al.*, “A connectome of the adult drosophila central brain,” *BioRxiv*, 2020.
7. M. Keramati and B. Gutkin, “A reinforcement learning theory for homeostatic regulation,” in *Advances in neural information processing systems*, pp. 82–90, 2011.
8. N. Yoshida, “Homeostatic agent for general environment,” *Journal of Artificial General Intelligence*, vol. 8, no. 1, pp. 1–22, 2017.
9. A. P. Badia, B. Piot, S. Kapturowski, P. Sprechmann, A. Vitvitskyi, Z. D. Guo, and C. Blundell, “Agent57: Outperforming the Atari Human Benchmark,” in *International Conference on Machine Learning*, pp. 507–517, PMLR, 2020.
10. M. Lanham, *Learn Unity ML-Agents-Fundamentals of Unity Machine Learning: Incorporate new powerful ML algorithms such as Deep Reinforcement Learning for games*. Packt Publishing Ltd, 2018.
11. P. Sunehag, G. Lever, S. Liu, J. Merel, N. Heess, J. Z. Leibo, E. Hughes, T. Eccles, and T. Graepel, “Reinforcement learning agents acquire flocking and symbiotic behaviour in simulated ecosystems,” in *Artificial life conference proceedings*, pp. 103–110, MIT Press, 2019.
12. J. Yamada, J. Shawe-Taylor, and Z. Fountas, “Evolution of a complex predator-prey ecosystem on large-scale multi-agent deep reinforcement learning,” in *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2020.
13. A. J. Lotka, *Elements of Physical Biology*, by Alfred J. Lotka. Williams & Wilkins, 1925.
14. V. Christensen and C. J. Walters, “Ecopath with ecosim: methods, capabilities and limitations,” *Ecological modelling*, vol. 172, no. 2-4, pp. 109–139, 2004.
15. A. M. Zador, “A critique of pure learning and what artificial neural networks can learn from animal brains,” *Nature communications*, vol. 10, no. 1, pp. 1–7, 2019.
16. I. Kotseruba and J. K. Tsotsos, “40 years of cognitive architectures: core cognitive abilities and practical applications,” *Artificial Intelligence Review*, vol. 53, no. 1, pp. 17–94, 2020.
17. S. W. Wilson, “The animat path to AI,” in *From animals to animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior* (J. A. Meyer and S. W. Wilson, eds.), pp. 15–21, MIT Press, 1991.
18. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.
19. T. Karlsson, “Multi-Agent Deep Reinforcement Learning in a Three-Species Predator-Prey Ecosystem,” Master’s thesis, Chalmers, 2021.
20. L. Seuront, *Copepods: Diversity, Habitat and Behavior*. Nova Science Publishers, 2014.
21. P. Ferrari and B. Kleve, “A Generic Model of Motivation in Artificial Animals Based on Reinforcement Learning,” Master’s thesis, Chalmers, 2021.
22. H. Glimmerfors and V. Skoglund, “Combining Reflexes and Reinforcement Learning in Evolving Ecosystems for Artificial Animals,” Master’s thesis, Chalmers, 2021.