# On the Mutual Influence of Human and Artificial Life: an Experimental Investigation

Stefano Furlan, Eric Medvet, Giorgia Nadizar, and Federico Pigozzi

Evolutionary Robotics and Artificial Life Lab, Department of Engineering and Architecture, University of Trieste, Italy
stefanofurlan6@gmail.com, emedvet@units.it, giorgia.nadizar@phd.units.it, federico.pigozzi@phd.units.it

## Abstract

Our modern world is teeming with non-biological agents, whose growing complexity brings them so close to living beings that they can be cataloged as artificial *creatures*, i.e., a form of Artificial Life (ALife). Ranging from disembodied intelligent agents to robots of conspicuous dimensions, all these artifacts are united by the fact that they are designed, built, and possibly trained by humans taking inspiration from natural elements. Hence, humans play a fundamental role in relation to ALife, both as creators and as final users, which calls attention to the need of studying the mutual influence of human and artificial life. Here we attempt an experimental investigation of the reciprocal effects of the human-ALife interaction. To this extent, we design an artificial world populated by life-like creatures, and resort to open-ended evolution to foster the creatures adaptation. We allow bidirectional communication between the system and humans, who can observe the artificial world and voluntarily choose to perform positive or negative actions towards the creatures populating it; those actions may have a short- or long-term impact on the artificial creatures. Our experimental results show that the creatures are capable of evolving under the influence of humans, even though the impact of the interaction remains uncertain. In addition, we find that ALife gives rise to disparate feelings in humans who interact with it, who are not always aware of the importance of their conduct.

## Introduction and related works

In the 1990s, the commercial craze of "Tamagotchi" (Clyde, 1998), a game where players nourish and care for virtual pets, swept through the world. Albeit naive, that game is a noteworthy instance of an Artificial Life (ALife) (Langton, 1997), i.e., a simulation of a living system, which does not exist in isolation, but in deep entanglement with human life. It also reveals that ALife is not completely detached from humans, who might need to rethink their role and responsibilities toward ALife. We already train artificial agents by reinforcement or supervision: trained agents are notoriously as biased as the datasets we feed them (Kasperkevic, 2015), and examples abound[1]. For instance, chatbot Tay shifted from lovely to toxic communication after a few hours of interaction with users of a social network (Hunt, 2016). The field of robotics is no exception to the case, and while robots, a relevant example of ALife agents, are becoming pervasive in our society, we—the creators—*define* and influence them (Pigozzi, 2022). One day in the future, a robot could browse for videos of the very first robots that were built, eager to learn more about its ancestors. Suppose a video shows up, displaying engineers that ruthlessly beat up and thrust a robot in the attempt of testing its resilience (Vincent, 2019). How brutal and condemnable would that act look to its electric eyes? Would our robotic brainchildren disown us and label us "a virus" as Agent Smith (the villain, himself an artificial creature) does in the "Matrix" movie (Wachowski et al., 1999)? At the same time, how would such responsibility affect the creators themselves?

Broadly speaking, when dealing with complex systems involving humans and artificial agents, whose actions are deeply intertwined, what results from the mutual interaction of humans and ALife? In particular, do artificial agents react to the actions of humans, displaying short-term adaptation in response to stimuli? Do these actions influence the inherited traits of artificial creatures, steering their evolutionary path and long-term adaptation? And, conversely, are humans aware of their influence on ALife? Do they shift their conduct accordingly?

We consider a system that addresses these questions in a minimalist way. We design and implement an artificial world (Figure 1), populated by virtual creatures that actively search for food, and expose it to a pool of volunteer participants in a human experiment. We consider three design objectives: (a) interaction, that is bidirectional between human and ALife; (b) adaptation, of creatures to external stimuli, including human presence; (c) realism, of creatures to look "familiar" and engaging for participants. Participants interact with the creatures through actions that are either "good" (placing food) or "bad" (eliminating a creature): we then record the participants' reactions. At the same time, creatures can sense human presence. We achieve long-term adaptation through artificial evolution, and, for the sake of realism, we design the creatures to be life-like. As a result, the goodness or badness of human actions can potentially

[1] `https://github.com/daviddao/awful-ai`

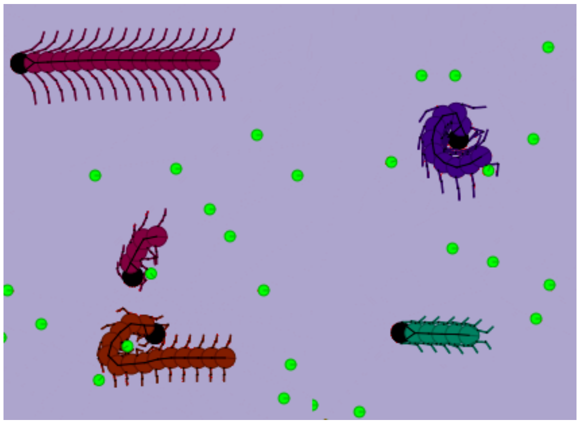


Figure 1: Our artificial world: worm-like agents are creatures that search for food (the green dots).

affect the evolutionary path of creatures, as well as their relationship with humans. Humans, on the other side, can feel emotions in the process. Participants thus play the role of a "superior being", absolute from any conditioning authority (Milgram, 1963), with power of life and death upon the creatures. Whether their actions will be good or bad is up to them: a philosophical debate on human nature that goes back to Thomas Hobbes (1651) and Jean-Jacques Rousseau (1755), with their opposing views propagating through history.

Other studies crafted artificial worlds, e.g., Tierra (Ray, 1992), PolyWorld (Yaeger et al., 1994), and Avida (Ofria and Wilke, 2004), with several different goals: they mostly investigate questions related to evolutionary biology (Lenski et al., 2003), ecology (Ventrella, 2005), open-ended evolution (Soros and Stanley, 2014), social learning (Bartoli et al., 2020), or are sources of entertainment and gaming (Dewdney, 1984; Grand and Cliff, 1998). Albeit fascinating, none of these addresses the main research question of this paper, i.e., the mutual influence of human life and ALife. Our work also differs from multi-agent platforms, whose focus is on optimizing multi-agent policies for a task (Suarez et al., 2019; Terry et al., 2021).

The work that is the most similar to ours pivots around the "Twitch Plays Robotics" platform of Bongard et al. (2018). While paving the way for crowdsourcing robotics experiments, it is, rather than an artificial world, an instance of "interactive evolution" (with participants issuing reinforcements to morphologically-evolving creatures), and does not detail the influence of creatures on participants.

We instead concentrate on the bidirectionality of interaction, and branch into two complementary studies: the first aimed at quantifying the effects of human interaction on artificial creatures, and the second focused on surveying how humans perceive and interface themselves with ALife. Concerning the former, we simulate human actions on the system and analyze the progress over time of some indexes, whereas for the latter we perform a user study involving a pool of volunteer participants interacting with the creatures. The experimental results confirm the importance of focusing on the bidirectionality of human-ALife interaction, and open a way towards more in depth analyses and studies in the field. Not surprisingly, we find that an artificial world subjected to human influence is capable of evolving, yet the real impact of human behavior on it, be it positive or negative, remains enigmatic. In addition, we discover two main currents of thought among people who interface themselves with ALife: those who feel involved and are aware of the consequences of their actions on an artificial world, and those who perceive ALife as a not attention-worthy far-fetched artifact.

## The artificial world

### Objectives

The aim of this work is to investigate the mutual influence of human life and ALife. We introduce an artificial world, populated by virtual creatures, that is suitable for such an investigation. We consider three objectives:

**Interaction.** In order to study any bidirectional impact between human life and ALife, the artificial world must support interaction. Moreover, interaction follows two design principles: (a) ergonomics, and (b) characterization. The former makes interaction easy and accessible for humans, while the latter is concerned with mapping an interaction back to the human behavior that generated it, and, in this study, classifying the interaction as either "good" or "bad". Last but not least, we remark that influence between human life and ALife must be bidirectional. Thus, for any human influence on ALife to happen, we require virtual creatures to be able to sense the presence of a human observer.

**Adaptation.** Second, we require the creatures inhabiting the artificial world to have the potential for adaptation to the environment and over time. During their life, virtual creatures undergo exposure to a set of stimuli, both "endogenous" and "exogenous": the former arise from the simulation itself (e.g., presence of food), while the latter arise from interaction with humans (e.g., good or bad actions). In order to evaluate any impact of human life on ALife, creatures should show adaptation to those stimuli, both in the short- and in the long-term, the first being a form of action-reaction, and the latter involving the development of more favorable traits.

**Realism.** To incentive interaction, humans should be able to relate with familiar entities. Creatures should then have a realistic look, possibly resembling natural organisms. In particular, they should have an appearance of "life" and engage in life-like activities, in order to elicit any notion of ALife in the observers. At the same time, creatures should not be too realistic, or even human-like, to avoid the notorious "uncanny valley" problem witnessed with highly-realistic robots (i.e., uneasiness and revulsion in the observers) (Mori et al., 2012).

## Environment and creatures

The artificial world introduced in this paper is visually two-dimensional, simulated in discrete time and continuous space, enclosed within an impassable rectangle of size $420 \times 240$ m. The colorful virtual creatures that populate it actively explore the space and hunt for food units lying on the ground (Figure 1). Each creature is endowed with a certain amount of energy, which dissipates at every time step and replenishes once the creature eats food. If a creatures depletes all of its energy, it dies and is removed from the world; then, a new creature is born by mutating one of the surviving creatures. Human observers may interact with the creatures by nourishing them, i.e., placing food in their proximity, (a "good" action) or eliminating some of them (a "bad" action).

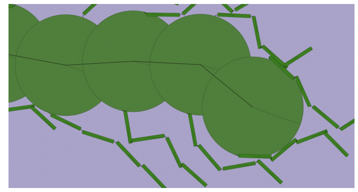
We implemented the project in the Java programming language, building on top of the dyn4j[2] physics engine, for which we set the time step to $\Delta t = \frac{1}{60}$ s, no gravity, and a linear speed damping coefficient that makes creatures movement appear like happening in a fluid. We made the project publicly available at `https://gitlab.com/step.lumumba/worm_simulator`.

We represent each creature as a *genotype*, that we map to a *phenotype*, i.e., the body and the brain of the creature.

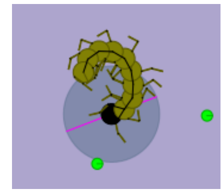
**Creature body.** The creatures have a worm-like body consists of a variable number of *segments* chained together, starting from a "head" segment. Each segment is a circular mass of weight $10\,\mathrm{kg}$ and radius $1.5\,\mathrm{m}$ and is connected to up to two other segments with a joint that allows for some rotation: as a result, the body can bend and appears flexible. Two flagella, implemented as flexible strings of tiny rectangles, extrude from each segment (see Figure 2a for a close-up) and have a sensory function (detailed below) and an aesthetic function. The genetic encoding of a body is a numerical vector $\boldsymbol{g}_{\text{morph}} \in \mathbb{R}^5$, which encodes the following phenotypic traits: the number of segments, the number of rectangles per flagellum, the length of rectangles of the flagella, the body color, and the length of sensory memory. We include color as a trait because it is neutral with respect to selection and survival, and allows us to verify there is no bias dictated by the representation or the evolution. For the length of sensory memory, see next sub-section.

To ensure body traits lie in meaningful intervals (e.g., flagella do not disappear), we map each gene $g_i$ to the corresponding phenotypic trait as $p_i = \frac{1+g'_i}{2}(p_i^{\text{max}} - p_i^{\text{min}}) + p_i^{\text{min}}$,

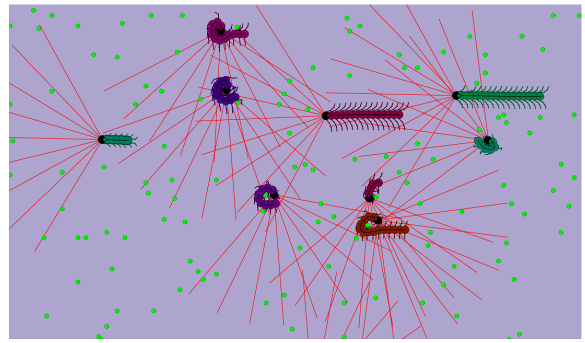
[2] `https://dyn4j.org`



(a) Close-up of a virtual creature.



(b) Creature with smell sensors.



(c) Creatures casting proximity sensor rays (depicted in red).

Figure 2: Details of creatures populating our artificial world.

where $g'_i = \min(\max(g_i, -1), 1)$: the first operand in the multiplication ensures the value lies in $[0, 1]$, and we then linearly rescale it to fit the interval $[p_i^{\text{min}}, p_i^{\text{max}}]$. After preliminary experiments, we set the intervals to be $\{5, 6, \ldots, 20\}$, $\{2, 3, \ldots, 40\}$, $[0.8, 1.8]$, and $\{0, 1, \ldots, 9\}$ for number of segments, number of rectangles per flagellum, length of rectangles of the flagella, and color, respectively. For color, integers in $\{0, 1, \ldots, 9\}$ correspond to $10$ possible colors.

By virtue of such morphological representation, creatures are in effect "primitive" enough to dispense with unnecessary complexity and focus on the mutual influence of human life and ALife; indeed, Mahoor et al. (2017) reported that the more "intuitive" the morphology, the more engaged participants to crowd-sourced robotics experiments are. Moreover, creatures do indeed recall natural organisms, in particular invertebrates (e.g., annelids, whose body consists of multiple segments), some of the simplest, most common, and most widely studied animals on Earth (Stewart, 2005). Our artificial world thus satisfies the Realism objective.

**Creature sensing.** We equip every creature with proximity, smell, touch, energy, temperature, and human presence sensors. Proximity sensors, depicted in Figure 2c, cast $9$ rays from the head circle and return the (normalized) distance from the closest object (either food or creature), clipping it to $1$ if there is none. Two smell sensors perceive the number of food units (over the total) in the right and left semi-circumferences of radius $9\,\mathrm{m}$ centered on the head, as shown in Figure 2b. Three touch sensors per side perceive whether one of three objects among food, other creatures, and the creature itself touch any of the flagella for that side,

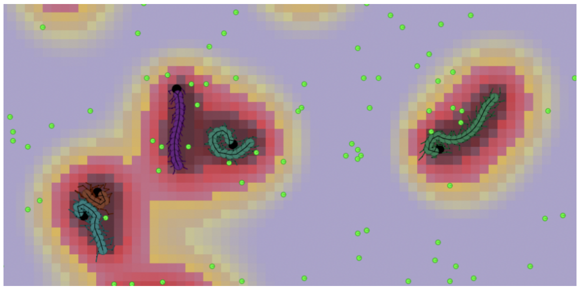


Figure 3: How creatures affect temperature, depicted as shades of red at every empty space (the darker, the warmer). Similarly, human observers condition temperature by moving their face across the screen.

and return 1 if yes, 0 if not. Energy sensor perceives the current energy of the creature, rescaled in $[0, 1]$. Temperature and human presence sensors, described below in detail, also return a value in $[0, 1]$ each. For every sensor reading, we also compute its trend over a window stretching $T$ time steps into the past, where $T$ is the fifth (and last) morphological gene and is thus subject to evolution. In this way, creatures can evolve some form of memory. Overall, creatures sense the world through $2(9+2+3\cdot2+1+1+1) = 40$ values in $[0, 1]$ at each time step.

Temperature and human presence sensors relate to the central piece of this study, and deserve an in-depth treatment. Temperature records the presence of "living" entities, either human or artificial, in the surrounding of a creature. For a simulation, we define the temperature matrix $\boldsymbol{H} \in \mathbb{R}^{+420\times240}$, where 420 and 240 are the side lengths of the artificial world. At each time step $k$ of simulation, we increment every $h_{x,y} \in \boldsymbol{H}$ by $\tau$ if there is at least one body segment whose center of mass lies within it. Then, we diffuse temperature by averaging it over the nine neighboring cells and multiplying by a damping coefficient $\alpha$: $h_{x,y}^{(k+1)} = \alpha\frac{\sum_{x',y'\in\{-1,0,1\}} h_{x+x',y+y'}^{(k)}}{9}$. Figure 3 is an instance of how creatures affect temperature during a simulation.

Moreover, humans (if present) do affect temperature. If a human observes the simulation on a computer screen, a computer vision system captures an image from the webcam placed over the screen and draws a bounding box around their face: we then increment $h_{x_c,y_c}$ by the area of the bounding box, $(x_c, y_c)$ being the coordinates of the center of the bounding box rescaled considering the world and captured image sizes. We remark that the area of the bounding box enclosing the observer's face, hence the temperature increase, depends on the proximity of the human to the computer screen. For a given creature, the temperature sensor perceives the average of the sub-matrix $\boldsymbol{H}_{\text{temp}} \in \mathbb{R}^{m\times m}$ centered on the head of the creature. After preliminary experiments, we set $\tau = 150$, $\alpha = 0.99$, and $m = 13$. We implemented face detection with the OpenCV library (Bradski, 2000), using Haar Cascades (Viola and Jones, 2001) as face detection algorithm.

Finally, the human presence sensor returns 1 if a human is observing the simulation, i.e., if the a face is detected by the webcam, and 0 otherwise.

By virtue of temperature and human presence sensors, creatures can sense the presence and location of humans, and thus affect one direction of the Interaction objective. Along the same direction (i.e., from humans to creatures), humans can influence the artificial world by placing food or killing creatures with a mouse click, as we shall see in Section "RQ2: human attitude towards ALife". In the other direction (i.e., from creatures to humans), the possible source of influence stays in the artificial world being depicted on the screen and hence being observable by humans.

**Creature brain.** We feed sensor readings and their trends to a feed-forward, fully-connected neural network with 40 input neurons (one for every sensor reading) and 3 output neurons, that correspond to the three possible actions for a creature: move ahead, to the right, or to the left. At every time step, we select the output having the highest absolute value and apply it as a force in that direction to the head circle. After preliminary experiments, we set one hidden layer with 10 neurons and $\tanh$ as activation function for all neurons. The genetic encoding for the controller is thus a numerical vector $\boldsymbol{g}_{\text{ctrl}} \in \mathbb{R}^{443}$ encoding the parameters of the neural network.

The genotype of a creature is then the concatenation $\boldsymbol{g} = [\boldsymbol{g}_{\text{morph}}\ \boldsymbol{g}_{\text{ctrl}}] \in \mathbb{R}^{5+443=448}$. Evolution operates on the representation $\boldsymbol{g}$, in a way that we detail in the next sub-section.

## Simulation

Simulation takes place in discrete time and continuous space. At every time step, $n_{\text{agents}}$ creatures and $n_{\text{food}}$ food units populate the artificial world.

At the very beginning, we initialize $n_{\text{agents}}$ creatures by sampling genotypes from $[-1, 1]^{448}$, i.e., each gene $g_i \sim U(-1, 1)$, mapping to the corresponding phenotypes, and giving birth to creatures at random positions, while making sure none of them overlap. At birth, we endow every creature with $e_{\text{init}}$ units of energy and set its generation to 0.

Then, at every time step of the simulation loop proceeds as follows:

1. Each creature senses the environment and uses the brain for processing sensor readings and producing an action.

2. The physics engine steps by applying the forces corresponding to each creature's action.

3. For each creature, if its head overlaps with a food unit, its energy is incremented by $e_{\text{food}}$ units. Upon the food consumption, the eaten food unit is removed from the world and a new one spawns at a random position.

4. For each creature, the energy is decreased by $e_{\text{step}}$ units. If energy of a creature drops to $0$, the creature dies and is removed from the world. As many food units as the number of its body segments spawn at the creature last position; to ensure a constant supply of food in the world, as many food units are randomly removed from the world.

5. For every creature just dead, if any, a new creature is born at a random position (making sure there is no overlapping), and its energy is set to $e_{\text{init}}$. With probability $p$, we randomly initialize its genotype by sampling $[-1, 1]^{448}$, and set the generation to 0; with probability $1 - p$ we perturb a parent genotype with Gaussian noise $\mathcal{N}(0, \sigma^2)$, to obtain a mutated copy of it, and set the generation to that of the parent plus one. In the latter case, we select a parent by performing roulette wheel selection (De Jong, 2016) on the age (i.e., number of time steps elapsed from birth) of the creatures. In this way, we use age as a proxy for fitness in our open-ended world, and ensure that the individuals most effective at surviving reproduce the most, while keeping some diversity in the population by choosing $p > 0$.

6. In the case of a human observer, they may interact with the creatures by performing "good" (placing food) or "bad" (eliminating a creature) actions, as we shall see in Section "RQ2: human attitude towards ALife".

By virtue of this procedure, our artificial world satisfies the basic conditions for evolution: selection of the fittest, variation of the offspring, and heredity (Darwin, 2004; Lewontin, 1970). Remarkably, evolution is a well-known example of an adaptation mechanism (Sipper et al., 1997): creatures must evolve to changes in their stimuli, including—in our case—human presence, leading us to satisfy the Adaptation objective. We remark that, as a consequence of the above procedure, both the number of creatures and of food units remain constant. In this way, we prevent the population from experiencing extinction before any interaction with humans and subsequent adaptation have taken place. After preliminary experiments, we set $e_{\text{init}} = 100$, $e_{\text{food}} = 20$, $e_{\text{food}} = 0.03$, $\sigma^2 = 0.35$, and $p = 0.1$.

## Experiments and discussion

We are interested in characterizing the mutual influence of human and artificial life. To this end, we performed an experimental evaluation and a user study aimed at answering the following two research questions:

RQ1 Does an artificial world subjected to human interaction evolve differently than without human interaction? Have "good" and "bad" human actions a different impact on the evolution?

RQ2 What is the attitude of humans towards ALife? In other words, are they aware of their influence on artificial systems? If so, do they change their behavior accordingly?

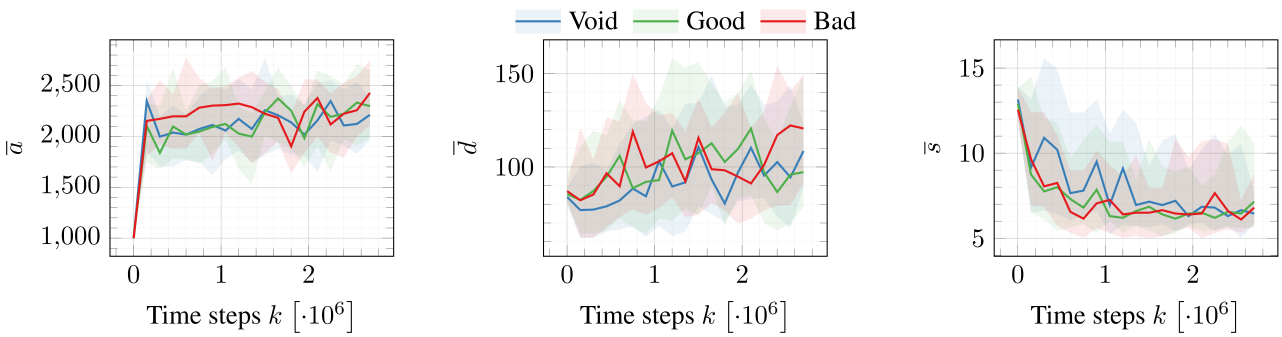
For addressing RQ1, we let our artificial world evolve under the influence of humans, i.e., with humans performing actions on it, and in the void, i.e., without human interactions. To evaluate the changes in the system, we took into consideration some indexes targeted at capturing variations in the artificial creatures. To make the human interaction long enough to impact on the evolution of our artificial creatures, we made use of simulated humans, displaying either good or bad behaviors.

Concerning RQ2, we designed a user study, with a pool of volunteer participants that interact with the simulator. In this case, we focused on appraising the attitude of humans towards our artificial world, by interviewing them and by examining the types of actions they conducted.

### RQ1: ALife evolution under human influence

To answer to RQ1, we performed an experimental campaign comprising three types of simulations. First we considered an in-the-void simulation (Void), employed as a baseline, where the virtual creatures are not subject to any exogenous stimulus, i.e., there is no human interaction. For the other two types of simulations, instead, we focused on estimating the impact of humans on the evolution of the artificial world. To this extent, we simulated human interaction with the artificial world at regular intervals during its evolution, to assess if such interactions could steer the evolutionary path of the system. We performed two variants of human-influenced simulations, the first comprising only "good" simulated humans (Good), and the second involving only "bad" simulated humans (Bad). We outline both variants in more detail in the next paragraph. For each type of simulation, namely Void, Good, and Bad, we let the system open-endedly evolve for approximately $2.4 \cdot 10^6$ time steps (corresponding to approximately 3000 generations in the Void case). In all simulations, we set $n_{\text{agents}} = 10$ and $n_{\text{food}} = 35$. For every type of simulation, we performed 30 independent runs, i.e., based on different random seeds, for a total of $3 \cdot 30 = 90$ runs.

For the experiments involving human intervention on the artificial world, i.e., Good and Bad, we simulate humans by replicating the aspects that characterize their interaction with the system. As described in Section "The artificial world", the influence of humans is twofold, unraveling into a temperature increase and in the possibility to perform active actions on the artificial world. To accurately capture both aspects, we repeat the following cycle every $20\,000$ time steps: (1) we mimic a human approaching the artificial world by increasing the temperature at a randomly chosen point of the world by $\Delta_\tau = 50\,000$ for $1000$ time steps (a new point being selected at each time step), and (2) we perform some actions, trying to counterfeit the behavior of a person interacting with the creatures. The actions we simulate are different for Good and Bad, in order to emulate the activity of




(a) Average age $\overline{a}$.

(b) Average traveled distance in the last 100 time steps $\overline{d}$.

(c) Average size $\overline{s}$.

Figure 4: Median and interquartile ranges of three indexes: age $\overline{a}$, traveled distance in the last 100 time steps $\overline{d}$, and size $\overline{s}$, averaged across the population. We use a different color for each type of simulation (Void, Good, or Bad).

a stereotypically "good" or "bad" human. In particular, we deem feeding creatures, i.e., placing food units near the head of a creature, as a positive action, hence associated to Good, while we consider killing a creature as a negative action, thus performed only in Bad. In both cases, we randomly select $r_{\text{agents}}$ creatures to undergo the chosen action, i.e., the feeding or the killing, with $r_{\text{agents}}$ randomly sampled from $\{1, 2, 3\}$ for every action.

In order to evaluate if evolution is actually taking place in the artificial world, we consider some indexes, which should capture the main features of the creatures. First, we consider the creatures age $a$, i.e., the amount of time steps since their birth, which should capture how well they adapted to survive in their environment. Then, to estimate movement, i.e., how lively creatures are, we examine the distance $d$ traveled by a creature in the last 100 time steps, instead of the total distance as this indicator would be strongly polluted by how long a creature survives. Last, we take into account the number of body segments composing a creature, i.e., its size $s$, as a morphological indicator to enlighten us on which agents features are more favored by evolution. For each index, we computed its average value across the living creatures every $150\,000$ simulation time steps, yielding the averaged indexes $\overline{a}$, $\overline{d}$, and $\overline{s}$. We report the median and the interquartile range of the aforementioned indexes throughout the runs in Figure 4.

From observing the plots of Figure 4, the answer to RQ1 is that evolution indeed takes place in all three types of simulation, steering the system in a clear direction for the three indexes. Hence, we can affirm that the considered system is suitable for studying the mutual influence of artificial creatures and humans.

Focusing on each subplot, we can gain more insight into different aspects of what is happening in the system. First, concerning the average age $\overline{a}$ of the creatures, displayed in Figure 4a, it appears to be rising with the progress of the simulation. Thus, we can conclude that artificial creatures are adapting to the environment, improving their survival rate by becoming more skillful. However, it is unclear if the creatures live longer because they have developed the trait of hunting, i.e., moving to target food, or if they are just randomly roaming the artificial world, thus maximizing the likelihood of encountering a food unit. Figure 4b does not show any apparent trend in this sense to support any of the two hypotheses. Last, we can reason on the morphological traits favored by evolution, by looking at Figure 4c. From the plot it is not difficult to notice how creatures become smaller as evolution progresses, as it is likely easier for them to move, hence increasing the probability of coming across food units.

To deepen our analysis and estimate the impact of human actions on ALife, we can study Figure 4 comparing the trends corresponding to Void, Good, and Bad. Since none of the plots shows significant differences among the colored lines, we assume that none of the measured indexes is impacted by human actions. In addition, the analysis of other indicators, e.g., the length of flagella or the area covered by creatures, here omitted for brevity, gave similar results as the ones of Figure 4. However, we are cautious on declaring that human actions do not affect ALife. In fact, the absence of tangible outcomes could be caused by the too few simulated human interventions on the system or by the random selection of creatures to undergo the chosen actions.

## RQ2: human attitude towards ALife

For providing an answer to the second research question, we moved our focus away from the system, to concentrate on the impact interacting with an artificial world has on humans. To this extent, we performed a user study involving of 36 unpaid volunteer participants, 12 females and 24 males, ranging from 18 to 57 years old, who were made to interact with the artificial world for a limited time span, and whose mindset and perceptions were registered by the means of two

| Question | Answers |
|---|---|
| Do you think artificial life exists? | Yes (✓), I don't know (**?**), No (✗). |
| How will you behave towards the creatures in the simulation? | Positively (👍), I am still undecided (**?**), Negatively (👎). |
| Do you think artificial creatures can suffer? | Yes (✓), Maybe (**?**), No (✗). |

Table 1: Pre-interaction questionnaire.

| Question | Answers |
|---|---|
| How would you rate your perceived involvement? | 1, 2, 3, 4, 5. |
| How would you define your behavior towards the creatures in the simulation? | Very Positive (👍👍), Fairly Positive (👍), Fairly Negative (👎), Very Negative (👎👎). |
| Do you think you have hurt these creatures? | Yes, definitely (✓✓), Yes (✓), I don't know (**?**), Not really (✗), Definitely not (✗✗). |
| If you have killed any creature, why have you? | - |

Table 2: Post-interaction questionnaire.

questionnaires.

Concerning the human-system interaction, we aimed at two goals: (a) arousing participants interest towards the artificial world, and (b) maintaining fairness and consistency across evaluations. For achieving the first goal, instead of employing a newly generated artificial world for each participant, we let the system evolve in-the-void for $100\,000$ time steps before interfacing it with humans, with the aim of having lively creatures displaying engaging traits. To tackle the consistency objective, we saved the state of the system (and of all the creatures populating it) after the preliminary in-the-void evolution, and we restored it upon each external interaction, to ensure every person was seeing the artificial world from the same starting point. In addition, each participant was given the same amount of time to interact with the system, which we set to $5\,\mathrm{min}$. We remark that we did not request participants to perform actions or even to pay attention for the entire duration of the experiment: if they were not interested anymore they could just sit idle and avoid active communication with the artificial world.

Since the ultimate goal of this experimentation was to assess the human perception of ALife and the attitude of humans towards artificial creatures, we gave great importance to registering the actions people performed in the simulation, together with their mindset. For the first, we simply recorded every action a person effected on the artificial world, taking note of the time, the location, and the type of action, i.e., placing food or killing a creature. Concerning the latter, we interviewed the participants before and after the experiments, asking them to fill out two short questionnaires. The pre-interaction questionnaire, reported in Table 1, aimed at evaluating the general approach towards ALife, together with the expected behavior towards creatures populating an artificial world. Similarly, we designed the post-interaction questionnaire, described in Table 2, to capture the feelings after interacting with ALife and to let participants self-assess their conduct.

We display aggregations of the collected data in Figures 5 to 7. First, we correlated the results gathered from the first question of both questionnaires, i.e., pre-interaction "Do you think artificial life exists?" and post-interaction "How would you rate your perceived involvement?", obtaining the heat map of Figure 5. This is the first noteworthy result of our study: participants who believe in the existence of ALife tend to feel more involved when interacting with artificial creatures. In particular, we speculate that such people perceive the importance of their role and the influence of their actions on the artificial world, thus feeling more concerned with it and more prone to actively interact with artificial creatures.

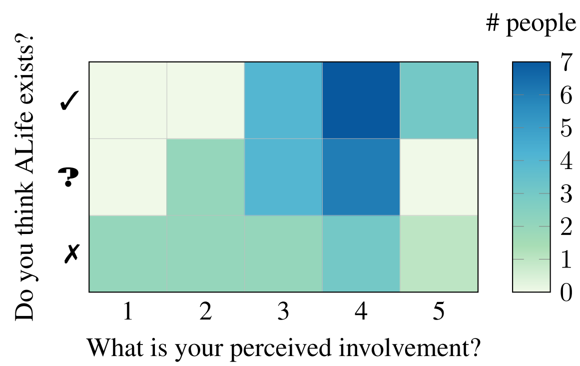



Figure 5: Relationship between participants views on the existence of ALife (pre-interaction "Do you think ALife exists?", on the $y$-axis) and their perceived involvement in the experiment (post-interaction "How would you rate your perceived involvement?", on the $x$-axis).

Moving on to Figure 6, we report the relationship between participants planned behavior (pre-interaction "How will you behave towards the creatures in the simulation?"), their self-assessed behavior (post-interaction "How would you define your behavior towards the creatures in the simulation?"), and the ratio of positive actions performed by each participant. The foremost observation we can make from such box plots, is that nobody decided to act negatively before interacting with the artificial world, which reveals a general tendency to avoid opting for negative actions in the first place. Such tendency is also confirmed by the overall ratio of good actions performed, which is always above 0.7. Focusing more on how participants rated their own conduct, we can notice that they are generally aware of the impact of

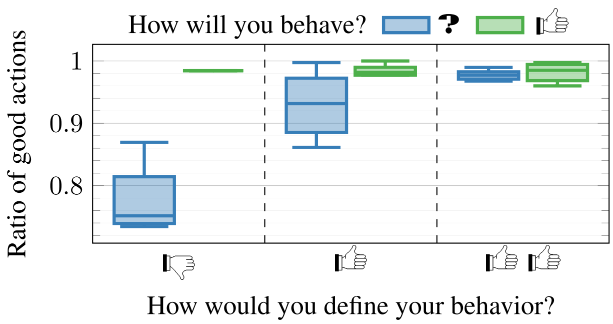



Figure 6: Relationship between participants planned behavior (“How will you behave towards the creatures in the simulation?”, color), self-assessed behavior (“How would you define your behavior towards the creatures in the simulation?”, $x$-axis), and the ratio of good actions performed ($y$-axis).

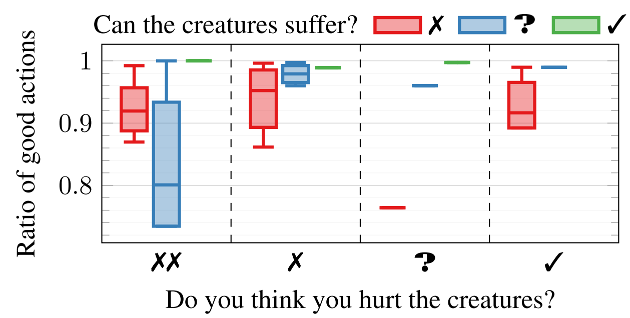



Figure 7: Relationship between participants feelings concerning the creatures ability to suffer (“Do you think artificial creatures can suffer?”, color), the perception about hurting them (“Do you think you have hurt these creatures?”, $x$-axis), and the ratio of good actions performed ($y$-axis).

their actions: those who described their behavior as fairly-negative (👎) show in general a lower ratio of good actions. Last, we can reason on participants coherence: in all cases those who decided to act positively towards the creatures in the artificial world (👍) show an averagely higher ratio of good actions performed than those who were undecided (**?**).

Another thought-provoking result is shown in Figure 7, where we report the results related to the participants perception of ALife suffering. From this figure, we note that the participants who perceive the creatures as alive, i.e., able to suffer, behave accordingly, trying to feed them and not to kill them. For the others, instead, the lower ratio of good actions performed reflects their perception of the creatures as a non-living artifact. At the same time, participants who performed more positive actions still believe they hurt the creatures far more than those who acted negatively. This, in our opinion, highlights how some people are extremely disconnected from ALife, and they consider it as a mere artifact. These results are in line with (Bongard and Anetsberger, 2016), which found that unpaid participants to a crowd-sourcing robotics experiment provided honest feedback.

Last, we examined the answers to the post-interaction question “If you have killed any creature, why have you?” to immerse ourselves in the motivations pushing participants to eliminate a creature from the artificial world. The collected responses were disparate, but they were mainly clustered into three categories: (a) curiosity, (b) mistakes, or (c) will to remove creatures displaying traits which people disliked. Among the listed categories, the first two ones were expected and, to us, reflect normal traits of human personalities. Conversely, the last answer is more disturbing.

Summing up, finding an answer to RQ2 is not straightforward. The experimental outcomes suggest that humans have mixed feelings with respect to ALife: some believe in its existence, feel involved when interacting with it, recognize their impact, and tend to act positively, while others perceive ALife as a pure fiction, not really worthy of attention and caution upon interaction.

## Concluding remarks

We studied the mutual influence between human and Artificial Life (ALife) from two symmetrical perspectives. First, we aimed at assessing the impact of an external superior entity on an artificial world, measuring if the evolutionary path of the system could be steered by human actions on it, and if the creatures populating it would adapt to the external influence, be it positive or negative. Furthermore, we tried to characterize the mindset and the behavior of people interacting with an artificial world where they played the role of superior entities yielding power of life and death upon its creatures.

To this end, we designed an artificial world based on the pillars of interaction, adaptation, and realism, and we performed a twofold experimental evaluation including real and simulated humans, focusing on the system evolution and on the human attitude.

Our results show that our artificial world is capable of evolving in the presence of an external influence, yet it is hard to appraise the impact of people on artificial creatures. Moreover, we find both positively involved cooperative attitudes and fairly detached negative perceptions with respect to ALife among participants. We believe our work stresses how delicate and contradictory is the relationship between the human and the artificial, the living and the machine. In the future, we plan to scale our experiment and carry on deeper analyses by involving more participants and encompassing expanded psychological and philosophical validations.

## Acknowledgments

The authors wish to thank the participants to the human experiment for their patience and dedication.

## References

Bartoli, A., Catto, M., De Lorenzo, A., Medvet, E., and Talamini, J. (2020). Mechanisms of Social Learning in Evolved Artificial Life. In *ALIFE 2020: The 2020 Conference on Artificial Life*, pages 190–198. MIT Press.

Bongard, J. and Anetsberger, J. (2016). Robots can ground crowd-proposed symbols by forming theories of group mind. In *ALIFE 2016, the Fifteenth International Conference on the Synthesis and Simulation of Living Systems*, pages 684–691. MIT Press.

Bongard, J. C., Cheney, N., Mahoor, Z., and Powers, J. P. (2018). The Role of Embodiment in Open-Ended Evolution. In *OOE3: The Third Workshop on Open-Ended Evolution*.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Clyde, A. (1998). Electronic pets. *Teacher Librarian*, 25(5):34.

Darwin, C. (2004). *On the origin of species, 1859*. Routledge.

De Jong, K. (2016). Evolutionary computation: a unified approach. In *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*, pages 185–199.

Dewdney, A. K. (1984). Core wars. *Sci. Amer.*

Grand, S. and Cliff, D. (1998). Creatures: Entertainment software agents with artificial life. *Autonomous Agents and Multi-Agent Systems*, 1(1):39–57.

Hobbes, T. (1651). *Leviathan or The Matter, Forme and Power of a Commonwealth Ecclesiasticall and Civil*.

Hunt, E. (2016). Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. *The Guardian*, 24(3):2016.

Kasperkevic, J. (2015). Google says sorry for racist auto-tag in photo app. *The Guardian*, 1:2015.

Langton, C. G. (1997). Artificial life: An overview.

Lenski, R. E., Ofria, C., Pennock, R. T., and Adami, C. (2003). The evolutionary origin of complex features. *Nature*, 423(6936):139–144.

Lewontin, R. C. (1970). The units of selection. *Annual review of ecology and systematics*, 1(1):1–18.

Mahoor, Z., Felag, J., and Bongard, J. (2017). Morphology dictates a robot's ability to ground crowd-proposed language. *arXiv preprint arXiv:1712.05881*.

Milgram, S. (1963). Behavioral study of obedience. *The Journal of abnormal and social psychology*, 67(4):371.

Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100.

Ofria, C. and Wilke, C. O. (2004). Avida: A software platform for research in computational evolutionary biology. *Artificial life*, 10(2):191–229.

Pigozzi, F. (2022). Robots: the century past and the century ahead.

Ray, T. S. (1992). Evolution, ecology and optimization of digital organisms. *Santa Fe*.

Rousseau, J.-J. (1755). *Discourse on the Origin and Basis of Inequality Among Men*.

Sipper, M., Sanchez, E., Mange, D., Tomassini, M., Pérez-Uribe, A., and Stauffer, A. (1997). A phylogenetic, ontogenetic, and epigenetic view of bio-inspired hardware systems. *IEEE Transactions on Evolutionary Computation*, 1(1):83–97.

Soros, L. and Stanley, K. (2014). Identifying necessary conditions for open-ended evolution through the artificial life world of chromaria. In *ALIFE 14: The Fourteenth International Conference on the Synthesis and Simulation of Living Systems*, pages 793–800. MIT Press.

Stewart, A. (2005). *The earth moved: on the remarkable achievements of earthworms*. Algonquin Books.

Suarez, J., Du, Y., Isola, P., and Mordatch, I. (2019). Neural mmo: A massively multiagent game environment for training and evaluating intelligent agents. *arXiv preprint arXiv:1903.00784*.

Terry, J. K., Black, B., Grammel, N., Jayakumar, M., Hari, A., Sullivan, R., Santos, L., Dieffendahl, C., Horsch, C., Perez-Vicente, R., et al. (2021). Pettingzoo: Gym for multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 34.

Ventrella, J. (2005). Genepool: Exploring the interaction between natural selection and sexual selection. In *Artificial life models in software*, pages 81–96. Springer.

Vincent, J. (2019). That video of a robot getting beaten is fake, but feeling sorry for machines is no joke.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. Ieee.

Wachowski, A., Wachowski, L., Reeves, K., Fishburne, L., Moss, C.-A., Weaving, H., Foster, G., Pantoliano, J., and Staenberg, Z. (1999). *Matrix*. Warner Home Video Burbank, CA.

Yaeger, L. et al. (1994). Computational genetics, physiology, metabolism, neural systems, learning, vision, and behavior or Poly World: Life in a new context. In *Santa Fe Institute Studies in the Science of Complexity*, volume 17.