Open-endedness and intrinsic motivation in embodied virtual agents

A Systematic Literature Review

Mircea-Tudor Lică[1]

Supervisor(s): Chirag Raman[1]**, Ojas Shirekar**[1]

[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 23, 2024

Name of the student: Mircea-Tudor Lică
Final project course: CSE3000 Research Project
Thesis committee: Chirag Raman, Ojas Shirekar, Willem-Paul Brinkman

Abstract

Virtual agents have demonstrated remarkable progress in both competitive and cooperative environments. Embodied agents, which enhance AI interactions with the physical world, show great promise for a variety of use cases in both virtual and non-virtual settings. This literature review examines the intersection of embodied virtual agents with cognitive and social frameworks derived from human behavior, focusing on open-ended learning and intrinsic motivation. These paradigms, inspired by human learning and adaptability, offer a path towards addressing the limitations of current artificial systems. The literature survey provides a thorough analysis of the research landscape, discussing the definitions, applications, and benefits of embodied agents in virtual settings. Furthermore, it evaluates the methods and benchmarks used to assess the capabilities of these systems, while offering possible solutions for developing the next generation of embodied agents.

1 Introduction

Following the emergence of deep learning in 2012 [1], virtual agents have experienced a significant increase in capabilities in both zero-sum games [2] [3] and cooperative settings [4] across multiple environments. Moreover, embodied agents represent a promising approach towards enhancing artificial intelligence (AI) interaction with the physical world, opening the path to numerous economically viable applications. One of the limiting factors in real-world usage of embodied agents powered by deep learning is the massive amounts of data required for training deep architectures [5]. Collecting data in real environments is often expensive and extremely time consuming, diminishing the usability and feasibility of any embodied agents by both researchers and practitioners.

Embodied virtual agents provide a possible solution to the data bottleneck by combining embodiment and virtual systems, operating in simulated environments where data collection is no longer an issue [6]. Furthermore, these methods are used to understand autonomous intelligent behaviour as a powerful synergy between the agent and the environment [7], providing an alternative view towards learning and intelligence compared to the static learning paradigm specific to supervised and self-supervised learning.

Despite the success of deep learning algorithms across a wide range of tasks, current AI systems lack the reasoning and generalization capabilities specific to humans [8], hinting towards possible limitations in the road to human-level AI [9]. For example, DreamerV3 [10] introduces an algorithm capable of performing tasks across a wide range of environments and types of agents. However, the model requires explicit reward signals from the environment and can only achieve mastery of a single task. To combat these issues and develop more autonomous and generalizable agents, scientists try to find inspiration in theoretical neuroscience by taking ideas from the inner workings of the human brain. This connection between neuroscience and AI has been beneficial for both fields, with famous examples like the perceptron [11], Convolutional Neural Networks (CNN) [12] or Dropout regularization [13].

The aim of this literature review is to explore how current research on embodied virtual agents leverages the cognitive and social frameworks derived from humans, with an emphasis on open-ended learning [14] and intrinsic motivation [15]. Open-ended learning is a learning paradigm where agents continuously adapt to new environments and tasks, without any predefined goals and termination conditions. On the other hand, intrinsic motivation deals with the ability of an agent to explore the environment without any external or explicit reward signals. These concepts stem from an animal's ability to learn and generalize to new environments with sparse or absent rewards. Thus, these two methods provide a promising approach towards the future of virtual and non-virtual agents, tackling the innate problems of the current systems [10].

In order to provide a comprehensive analysis on current literature, together with possible directions for future research, the survey will consider different viewpoints and highlight their limitations. More specifically, we cover various definitions of embodied agents, and explore how open-ended learning and intrinsic motivation have been applied in the context of embodied virtual agents. Moreover, we study the gap between virtual and non-virtual agents to quantify the advantages and disadvantages of constraining agents inside a virtual setting. Lastly, we analyze the literature corresponding to the current benchmarks and evaluation methods considered when assessing an embodied agent's capabilities.

Below we highlight the main questions we aim to answer throughout the literature review:

Question 1. *How has open-ended learning and intrinsic motivation been applied in the context of embodied virtual agents?*

Question 2. *What kind of benchmarks are used to assess an agent's open-endedness and intrinsic motivation capabilities?*

Question 3. *What are the limitations of the current methods in the field and what are the possible directions for future research?*

The structure of the survey is as follows. Section 2 provides a comprehensive analysis on the methodology used to filter the relevant literature based on the objectives of the survey. Section 3 provides the reader the necessary background on the topics presented in Section 4. In turn, Section 4 provides an overview on the current literature concerning open-ended learning and intrinsic motivation in the context of embodied virtual agents, while also considering the current benchmarks specific to these methods. Section 5 introduces a discussion based on the findings in the previous two sections, followed by Section 6 which proposes future research directions based on current trends in generalizable agents.

2 Methodology

In order to minimize potential conflicts or ambiguity in the selection of the scientific articles considered in this survey, we follow as close as possible the standardized Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) [16] guidelines to provide transparency and reproducibility to the reader. The following subsections describe in detail the retrieval, selection and inclusion criteria used to filter the papers considered in the study, enabling full reproducibility and transparency. Figure 1 provides an overview of the methodology, highlighting both the selection procedure and the results.

2.1 Literature search

Taking in consideration the aim of this survey and its intersection with multiple fields such as neuroscience, machine learning and robotics, we mainly consider scientific articles from multiple databases, encompassing different perspectives and viewpoints. With this objective in mind, we choose the following databases: IEEE Xplore [1], Scopus [2], Nature [3] and ACM Digital Library [4].

The aforementioned databases consist of millions of records, most of which are not in the scope of this literature review. Thus, we search for relevant articles using keywords, boolean operators and regex patterns. By using keywords related to the research questions highlighted in Section 1, we restrict our search exclusively to a small subset of articles, concentrated around the goal of the analysis. Moreover, using the boolean operators AND, NOT, OR, we construct more fine-grained search queries, reducing the solution domain even further. Figure 2 features the final search query used across all databases.

```
("intrinsic motivation*" OR "inherent
motivation*" OR "open-endedness" OR
"open-ended learning") AND ("embodied"
OR "embodiment" OR "robot*") AND
("agent*" OR "multi-agent*" OR "multi
agent*") AND NOT "psychology"
```

Figure 2: Search query used for database filtering. The relational operators are highlighted with blue.

[1]https://ieeexplore.ieee.org
[2]https://www.scopus.com
[3]https://www.nature.com
[4]https://dl.acm.org

2.2 Literature filtering

Once a search query is determined, the next procedure consists of filtering all the articles found based on the title and the abstract, effectively eliminating any duplicates or irrelevant articles found. Furthermore, we showcase the inclusion and exclusion criteria used to filter out the scientific articles during the full-text retrieval stage.

Inclusion criteria

- Journal article/Conference proceedings written in the English language
- Focus on intrinsic motivation or open-ended learning
- Focus on benchmarks or evaluations for embodied virtual agents

Exclusion criteria

- Full-study not available
- Journal article/Conference proceedings written in other language than English
- Other focus than open-ended learning/intrinsic motivation and emboied virtual agents

Lastly, to showcase an overview of the results from the literature search, Figure 3 dissects the list of articles, by grouping them into specific categories that directly relate to Question 1, 2 and 3.

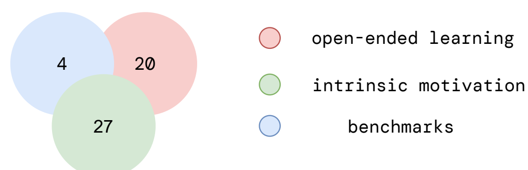

Figure 3: Number of papers covering each focus point. Each color highlights one of the three key concepts discussed throughout the literature review.

2.3 Citation chaining & Other sources

In order to broaden the scope of the articles considered after the literature filtering, we use citation chaining to select highly relevant papers that might have not been included in the original search due to the choice of keywords. Citation chaining, also known as backward chaining, considers the cited articles in the already selected papers as a proxy for their relevance to the subject and credibility. In the context of this literature review, citation chaining contributed with five papers to the final number of studies included for synthesis. Furthermore, we enhance the set of studies included in this paper with additional articles found on Google Scholar [5]. To encourage transparency, the complete list of articles from Google Scholar can be found in the Appendix A.

[5]https://scholar.google.com

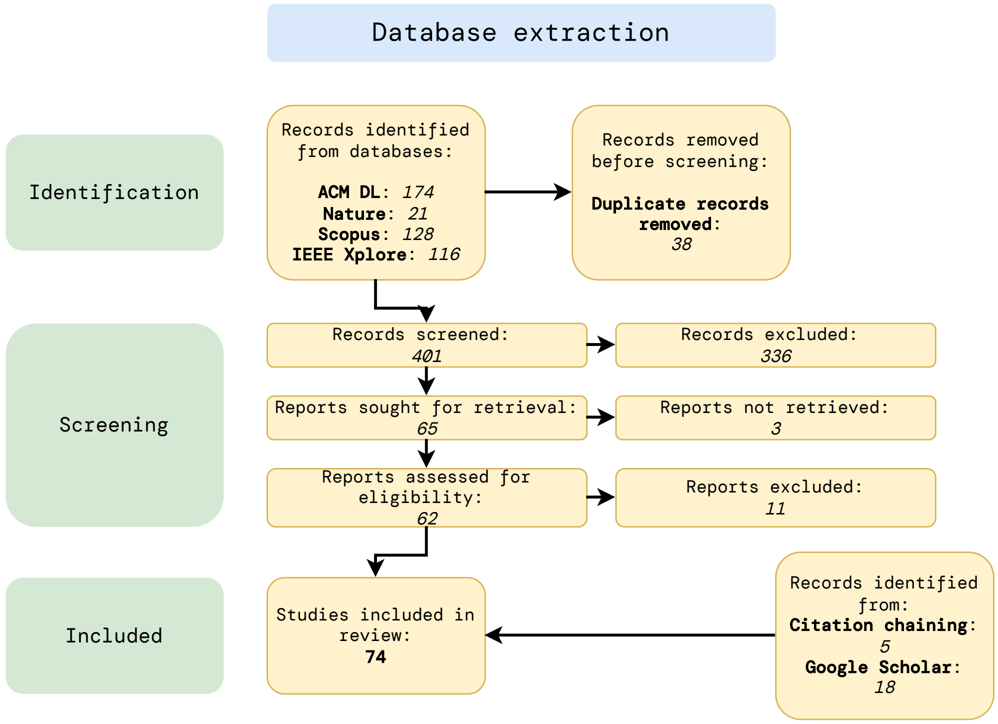

Figure 1: PRISMA diagram visualization. Provides a high-level overview of the literature search conducted for this survey. Sections highlighed with green correspond to different stages of the search.

3 Background

The purpose of this section is to introduce the reader to the core concepts related to the literature considered in this review. Specifically, each of the following three subsections briefly present the background and motivation behind these ideas, while providing a clear definition to assist the reader for the rest of the review.

3.1 Embodied virtual agent

For the scope of this survey, we define a virtual agent as a software-based system that operates in a virtual setting, capable of executing actions dependent on the state of the environment. For example, a primary use case of virtual agents are chatbot applications, where the agent responds according to a query formulated by a user.

Embodiment on the other hand, bridges the gap between virtual agents and physical reality by providing agents with means of perceiving the world through sensory and physical experiences. Thus, embodied virtual agents leverage the synthesis with the physical world to perform a wide range of tasks and real-world applications, such as navigation [17] [18], visual exploration [19] or embodied question answering [20].
The idea of embodied agents comes mostly from the concept of embodied cognition, which states that cognition requires acting with a physical body in an environment. Thus, understanding the underlying cognitive processes requires knowledge about their relationship with the agent's embodiment and the sensory signals coming from the environment. Furthermore, the concept of embodied cognition is also prevelant in social contexts [21], where non-verbal information such as body language or face expressions become a good indicator of the other's agent state.

Considering prior literature and different schools of thought in embodiment, for the purpose of this survey, we formally define an embodied virtual agent in Definition 1.

> **Definition 1.** *An embodied virtual agent is an autonomous intelligent agent that interacts with the environment through a form of physical embodiment that can perceive sensory inputs.*

3.2 Open-Endedness

The concept of an open-ended system has been widely discussed through the years, having plenty of proposed definitions [22] [23] [24] [25]. While their interpretation widely varies, they all share an overarching theme where an open-ended system endlessly produces novel and interesting artifacts with respect to an external observer.

The majority of formal definitions for open-endedness come in the context of evolutionary systems [24] [25] [22], where the complexity of the systems increases in time as a result of evolution. Subsequently, the steady increase in complexity is translated into novel behavioural artifacts. On the other hand, there is an increasing body of research, defining open-endedness in terms of an external observer [23] [26]. Compared to the evolutionary perspective, open-endedness is now seen through the "eyes" of the observer, which quantifies the increase in novelty of the system through time, based on its own measure of interestingness. Definition 2 provides a general yet quantifiable formulation for open-endedness in the context of virtual agents, which combines both schools of thought mentioned in the beginning.

Definition 2. *Open-endedness is the ability of an agent to improve and adapt over time, being able to constantly generate novel behaviour with respect to an external observer.*

3.3 Intrinsic motivation

Intrinsic motivation (IR) is closely linked with open-ended learning, and has its roots in psychology [27] [28], where it provides a framework for understanding animal behaviour in the absence of any major external stimuli [29]. This perspective becomes attractive from a computational perspective, by providing a way of inducing exploration and learning in virtual agents, without relying on predefined learning signals.

Besides psychology, intrinsic motivations are also present in theoretical neuroscience, where it is considered to be responsible for generating learning signals in the brain if the organism is acquiring new skills [30]. Thus, by combining the psychology and theoretical neuroscience research we formalize intrinsic motivation in Definition 3.

Definition 3. *Intrinsic motivations are mechanisms that drive learning of new skills and knowledge, without the need for any homeostatic rewards.*

As stated in the beginning of the section, open-ended learning and intrinsic motivation are intertwined concepts, that cannot exist in complete separation. This synergy allows agents to demonstrate continuous self-driven growth and adaptation.

4 Results

We now analyze the current literature landscape for both open-ended learning and intrinsic motivation. More specifically, we focus on scientific work which relates to the questions specified in Section 1, combining the two paradigms with the concept of embodied virtual agents. Thus, Section 4.1 summarizes scientific works related to open-endedness and open-ended learning, while Section 4.2 focuses on intrinsic motivation research. Finally, Section 4.3 dissects current available benchmarks that quantify motivations and open-endedness in embodied virtual agents.

4.1 Open-ended learning

Based on the methodology described in Section 2, we separate open-ended research in four categories: social learning, evolutionary methods and curriculum learning approaches. While this grouping is not disjoint, we choose this configuration in order to focus on accuately presenting the main paradigms in open-ended learning.

Social learning

An important area of research in open-ended learning involves multi-agent settings, where similarly capable actors communicate with each other to exhibit more complex behaviors [31] [32] [33] [34] [35]. Inspired by theories stating that social interaction is a necessary component of intelligent behaviour [36], social learning provides a promising path towards open-endedness, by leveraging cooperation between multiple embodied entities.

Current research on vanilla model-free or model-based RL methods do not naturally utilize social learning for acquiring new skills or improving existing ones [10] [37]. As highlighted by [34], in order for social learning to emerge, certain environment conditions and agent architectures are required. For example, [34] states that the characteristics of either the environment or the agent, can influence the reward structure. On the other hand, [32] draws inspiration from psychology and associates social learning as a teacher-student scenario, where a more capable agent aids a weaker actor to acquire a skill faster. Moreover, [31] finds that equipping social agents with a conversational memory can drastically improve goal attainment and learning efficiency.

Once social learning occurs, it enables agents to learn complex behaviours, ultimately surpassing the expert agents. This cooperative improvement can be seen as a variant of self-play or self-improvement, found in algorithms like AlphaGo [38] or MuZero [3]. With respect to open-endedness, social learning improves generalization across a multitude of tasks in the presence of experts, which can be represented either by humans or more capable agents. Moreover, the student benefits from the interaction with the expert even in its absence, thus managing to outperform individually trained agents.

Another important concept related to social learning is cultural transmission [35]. The research paper associate few-shot learning with the same phenomenon found in human culture, where humans accumulate and refine their skills across multiple generations, with only limited interactions. The agent manages to achieve real-time imitation of a human counterpart with only a handful of examples, showcasing the potential of the method.

While social learning can promote cooperation, it also increases the risk of deception where agents manipulate other entities into adopting policies not in their best interest, as shown by CICERO [39]. While CICERO does not involve any social learning, it provides an example of possible adversarial social learning scenarios, and its possible implications. In this case, expert agents can deliberately hamper the learning capabilities of other agents for personal or group incentives.

Evolutionary methods

As described in Section 3.2, open-ended learning systems are not constrained to a fixed set of goals or tasks, but can autonomously produce new behaviors, skills, or solutions over time without any constraints. Evolutionary algorithms [40], which simulate the processes of mutation, selection, and recombination found and studied in natural evolution, are a promising approach for realizing open-endedness in virtual and non-virtual agents [41] [42] [43] [44]. In this scenario, learning can emerge by iteratively generating novel variations while keeping desired attributes, leading to out-of-distribution behaviours and capabilities.

There are several evolutionary techniques being explored for open-endedness including novelty search [43] [45], minimal criterion coevolution [42], and quality diversity algorithms [46]. Novelty-based search works by rewarding novel behaviours rather than incentivizing progress towards a fixed goal. On the other hand, coevolution focuses on the interaction between two coevolving populations. At the same time, quality diversity algorithms aim to generate multiple near-optimal solutions rather than a single optimal candidate, optimizing for both performance and diversity in their candidate set.

While evolutionary algorithms are established in literature for a long time, the combination between evolutionary methods and open-ended settings represent a relatively new research direction. Thus, the careful combination of evolutionary operators with open-endedness would allow future agents to continuously adapt to new environments and tasks.

Curriculum learning

Lastly, we look at curriculum learning, a technique that trains a machine learning model by exposing it to increasingly more difficult training data [47] [48]. In case of embodied virtual agents, curriculum learning translates to a method which allows the agent to progressively learn to solve tasks of increasing complexity by interacting with an environment. Plenty of research has been conducted at the intersection of open-endedness and curriculum learning, showing promising results in complex environments [49] [50] [48].

Firstly, [49] proposes a domain-independent goal generation mechanism to generate goals at different levels of complexity. The method is validated using a virtual mobile robot, which manages to produce compound goals by combining previously experienced states. Moreover, [48] presents the Intrinsically Motivated Goal Exploration Processes (IMGEP) algorithm to mimic developmental learning inspired from children behaviour into machines. IMGEP aims to generate its own goals and explore the environment with an incremental goal policy augmented with information reuse for better generalization. Lastly, Voyager [50] showcases an open-ended embodied agent guided by Large Language Models (LLMs). Voyager uses automatic curriculum learning (ACL) to achieve impressive generalization and adaptability results in the Minecraft environment [51].

As highlighted by the articles presented above, the potential role of curriculum learning in open-ended learning is two fold. Firstly, by using a curriculum, the agent is only exposed to novel tasks that are within cognitive reach. Secondly, curriculum learning helps break complex tasks in sequential milestones, guiding the exploration of the agent.

4.2 Intrinsic motivation

Similar to Section 4.1, we categorize the scientific literature in intrinsic motivation based on how they quantify motivation and exploration. For the purpose of this survey, we focus on methods that leverage computation and can be employed by a virtual agent. Lastly, we take a look at the difference between intrinsic and extrinsic motivation, and provide an overview of interactional motivation.

Knowledge-based methods

Knowledge-based methods of intrinsic motivation can be classified in two categories: novelty-based and prediction-based models. Novelty-based models [52] perform learning by comparing new experiences with old ones, shaping the reward signal to match the novelty of the situation. For example, exploring an unknown section of a maze would result in a high intrinsic reward, while moving in the same static area would not give any positive reward signal, due to the lack of novelty. On the other hand, prediction-based models [53] use the prediction error as a guide for the intrinsic reward signal. The prediction error is calculated between predictions made by the agent and the true observations from the environment. In this context, a high prediction error corresponds to a new or difficult to predict situation which translates in a high reward signal, guiding the exploration of the agent.

An example of novelty-based intrinsic learning includes the operation of service robots [54]. The authors take inspiration from biology and explore the concept of habituation [6] as a novelty detector. Similarly, [55] proposes a new intrinsic motivation mechanism Group Intrinsic Curiosity Module (GICM) that encourages the agent to pursue novel situations. Furthermore, [56] presents a novel algorithm for intrinsic motivation based on boredom and chaos theory. Specifically, boredom is encoded within the model using a chaotic element that generates conditions for exploring routes with no reward. Finally, Continual Curiosity driven Skill Acquisition (CCSA) [57] avoids dealing with high-dimensional spaces, and manages to define novelty in a compact low-dimensional space.

Competence-based models

Compared to knowledge-based models which leverage novelty and surprise to provide intrinsic reward signals instantaneously, competence-based models measure the capabilities of the agent over extended periods of time [58] [59] [60]. The intrinsic motivation reward signal comes from a significant shift in performance for the agent, regardless if the error is decreasing or increasing.

GRAIL [60] is an embodied robotic architecture designed for open-ended goal-discovering. GRAIL uses intrinsic motivation based on competence-based models to autonomously drive learning towards easy or achievable goals, by leveraging

[6] revisiting past experiences with the hope of finding something new.

changes in the environment over a specific time window. By only tracking changes in a finite time window, the robot is intrinsically discouraged to pursue too complex or out of reach tasks. Furthermore, its newest variation, C-GRAIL (Context GRAIL)[61] is able to assign different values to different goals, depending on the context of the agent.

Another research direction related to competence-based motivations is goal-directed empowerment [59]. Empowerment is defined as an information-theoretical measure related to the capacity of the agent to influence its environment. Formally, empowerment can be defined as the maximum amount of information that an agent can transmit to its future sensory perception system. If the future consists of multiple possible states, corresponding to a high entropy system, then the agent is incentivized to continue the exploration.

Current research in intrinsic motivation largely focuses on competence-based models, demonstrating better performance than knowledge-based methods. In terms of embodiment, the majority of work leverages robot virtual environments to assess the capabilities of the agents.

Interactional motivation

The other side on the motivation spectrum that drive virtual and non-virtual agents is extrinsic motivation. As a concept, extrinsic motivation can be best understood in comparison with intrinsic motivation, where there is usually non-existant or episodic reward signals. In contrast, extrinsic motivation refers to the ability of the agent to be guided by a series of reward signals given by virtual or non-virtual environments. Extrinsic motivation is predominantely used in virtual settings, where reward signals are considerably easier to model.

Interactional motivation [62] provides a method to induce self-motivation in artificial agents, combining ideas from both intrinsic and extrinsic motivation. Formally, interactional motivation is a mechanism that associates a value function with possible interactions between the agent and the environment, using an unsupervised learning mechanism to learn to maximize the value function over time. The difference between interactional and intrinsic motivation relies in the fact that the value function is explicitly defined.

4.3 Benchmarks

The purpose of this section is to provide the reader a comprehensive analysis on benchmarks and environments commonly used to assess the performance of intrinsic motivation and open-ended approaches in embodied virtual agents. Benchmarks play an important role in scientific research by providing standardized metrics for evaluating and comparing the performance of different methods, models, and algorithms, while promoting transparency. We emphasize that Reinforcement Learning (RL) is known for its stochasticity during evaluations [63], due to the finite number of training runs reported.

Open-ended learning and intrinsic motivation research spans a wide number of embodied environments and benchmarks [64] [65] [66] [67] [68] [69] [70] [71]. These studies demonstrate the broad applicability as well as the maturity of the field towards generalizable agents. For instance, [64] uses the Poppy humanoid robot to develop an active learning architecture that learns the most efficient data collection technique, leveraging intrinsic motivation algorithms.

Shifting towards virtual environments, [65] proposes the Open-ended Physics Environment (OPEn) to benchmark the ability of intelligent agents to perform downstream physical reasoning tasks using their model of the world, which translates in the open-endedness capabilities of the agent. Besides proposing the benchmark, the authors find that all the models fail in terms of sample efficiency on any downstream tasks, uncovering the limitations of the current algorithms. Furthermore, [66] introduces the Robot open-Ended Autonomous Learning (REAL) benchmark to promote open-ended learning research in developmental robotics. The benchmark consists of 2 phases, the first being skill acquisition through open-ended learning and intrinsic motivation in a virtual robot-arm environment, while the second one focuses on testing the capabilities of the robots on unknown tasks. On the other hand, [67] and NeuralMMO [68] propose multi-agent environments for both collaboration and competition scenarios. Specifically, [67] focuses on multi-agent continuous control, while NeuralMMO focuses on multi-agent competition in a finite resource environment. In terms of embodiment, NeuralMMO uses a three dimensional animated character and a toolbox, such that the agent can successfully interact with the environment. Another popular setting for testing embodied virtual agent is Minecraft [69]. Due to the design of the environment, as well as the incremental difficulty of tasks an agent needs to perform, Minecarft represents an ideal setting for open-ended research.

Lastly, we look at a relatively new benchmark for open-endedness in minimal criterion coevolution [70], described in Section 4.1. The paper introduces a maze environment that allows for infinite expansion in size and complexity, ideal for open-ended learning. The mazes have a two dimensional structure and are procedurally generated.

5 Discussion

We now analyze the information extracted in Section 4 and come up with a series of findings we believe are representative for the scope of the literature review. The goal is to bring into attention current advantages and limitations in order to guide the reader in the right direction.

One of the key insights drawn from the reviewed literature is the importance of social multi-agent cooperation as highlighted in Finding 1. Leveraging cooperation between agents can result in a more sample efficient convergence towards completing complex tasks and adaptation to new scenarios. Moreover, social interaction and cooperation can result in emergent behaviour, exceeding the initial capabilities of the agents.

Finding 1. *Social interaction and cooperation provide a mechanism of recursive collaborative improvement, enhancing adaptability and efficiency.*

While mostly overlooked by current literature, the choice of embodiment significantly alters the capabilities of the agent. This phenomenon is present in both single and multi-agent setting, having implications in collaborative settings and in the agent-environment interaction. More specifically, less embodiment capabilities results in fewer communication channels and fewer ways to interact with the virtual environment, as stated in Finding 2. In social environments, the correct choice of embodiment can convey more information than standard communication, exploiting non-verbal information such as posture, facial or body expressions.

Finding 2. *Agent's embodiment choice directly influences its synergy with the environment, and possibly hampering its adaptability.*

The Sim2Real gap, a significant challenge in the deployment of embodied virtual agents to the real world, refers to the discrepancy between simulations and real-world environments. The majority of virtual agents trained in simulated environments tend to underperform when transferred to real-world settings, due to the inherent complexity of reality. In the context of open-ended learning, the Sim2Real gap becomes more evident, as current methods in open-ended learning require large amount of training data and tasks. Currently, these requirements are rarely met when deploying an agent in the real world, limiting the usability of open-endedness in artificial agents. Thus, following the current line of research, we highlight the need for more sample efficient solutions to open-ended learning in Finding 3.

Finding 3. *Deploying open-ended agents in the real world, requires the development of more sample efficient learning algorithms.*

Another limitation of the current research landscape is the absence of established benchmarks and environments for testing the capabilities of embodied virtual agents. While Section 4.3 covers a multitude of existing environments, there is almost no intersection between experiments conducted and published by different research groups. Without clear consensus on standard de-facto benchmarks, similar to ImageNet [72] or COCO [73] in Computer Vision, tracking progress in the field becomes increasingly harder. This leads us to Finding 4.

Finding 4. *Both open-ended learning and intrinsic motivation research suffer from the lack of standardized benchmarks.*

Lastly, we highlight the characteristics of open-ended learning and intrinsic motivation and their importance towards achieving generally capable and intelligent agents. Firstly, intrinsic motivation in an agent enables autonomous behaviour, favoring unbounded exploration and improvement. Secondly, open-ended learning enables the generalization and adaptability required by agents in the real world. Thus, we advocate for the necessity of both open-endedness and intrinsic motivations for creating generally capable agents, as highlighted in Finding 5.

Finding 5. *Open-endedness and intrinsic motivation are necessary for emergence of generally capable agents.*

6 Future directions

By examining the limitations of the current research landscape described in Section 5, we highlight what we believe to be the most promising directions for the next generation of embodied virtual agents.

The Bitter Lesson [7] states that methods that leverage computation most effectively usually perform the best. Thus, given the current state of the art paradigms in deep learning, the quality of the models is expected to increase over time. Thus, large pretrained models such as Large Language Models [74], Vision-Language Models [75] or World Simulators [76] could theoretically solve both the data bottleneck problem as well as the Sim2Real gap.

There is an increasing body of research for leveraging the general knowledge of large pretrained models to guide the embodied agent through both learning and exploration [50]. Moreover, recent work like Genie [77] could solve the data required for open-ended learning by leveraging a world simulator, capable of generating training data for the agent. Besides generating additional training data, these models can also be used to guide the agent using preferences [78]. For instance, Vision-Language Models could be used to assist the embodied agent in choosing between two different trajectories based on some predefined goal. Furthermore, natural language could be used to increase the efficiency of deep RL [79] by steering the agent in the right direction. Lastly, the Sim2Real gap might disappear completely once pretrained models are used for low-level robotic control. Recent work like RT-2 [80] already hints towards the advantages of using Vision-Language Models for robotics.

7 Conclusion

In conclusion, across this literature review we highlighted the significant advancements and ongoing challenges in the field of embodied virtual agents, focusing on open-ended learning and intrinsic motivation techniques. We emphasize how the integration of different cognitive and social frameworks inspired by human behaviour provide promising results towards a new generation of embodied agents. Furthermore, we underlined the benchmarks and environments used to assess the capabilities and shortcomings of these algorithms, where we find a lack of standards which could potentially create confusion in the research community.

Despite the notable progress, there are significant bottlenecks towards achieving autonomous and generalizable AI systems

[7] http://www.incompleteideas.net/IncIdeas/BitterLesson.html

in both virtual and non-virtual settings. The data and computational resources required by current methods make the transfer between virtual and non-virtual environment impossible. To this end, we proposed the usage of large pretrained models to bridge the generalization gap and increase sample efficiency. Overall, this review underscores the importance of continued exploration into open-ended learning and intrinsic motivation to push the boundaries of what embodied virtual agents can achieve, moving us closer to the goal of creating truly intelligent and adaptable AI systems.

8 Limitations

In this section we highlight the main limitations of the survey considering the timeline of the project and paper constraints. While the literature review covers the main topics in both open-ended learning and intrinsic motivation, some areas are partially or completely overlooked. Specifically, by focusing on computational methods, we do not fully cover the literature on open-ended learning and intrinsic motivation in psychology and neuroscience. Moreover, we only briefly mention about the importance of embodiment in the performance and adaptability of a virtual agent, which turns out to be a crucial factor especially in social and multi-agent settings. Despite these shortcomings, the literature review contains enough information about these topics to guide the reader for a more in-depth analysis. Finally, due to space constraints, we do not have the necessary space to provide a mathematical background to the algorithms related to intrinsic motivation and open-ended learning.

9 Responsible Research

Responsible research, especially in the context of a literature review, involves the meticulous and ethical synthesis of existing knowledge, ensuring accuracy, reliability, and integrity throughout the process. To this end, we made all possible efforts to make the methodology as transparent as possible for full reproducibility. Moreover, we follow the established PRISMA workflow to ensure the selection process and filtering decisions are well understood. Finally, in Appendix B we fully disclose the usage of Large Language Models to perform this literature review.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[2] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, "Mastering the game of go with deep neural networks and tree search," *nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[3] J. Schrittwieser, I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt, A. Guez, E. Lockhart, D. Hassabis, T. Graepel *et al.*, "Mastering atari, go, chess and shogi by planning with a learned model," *Nature*, vol. 588, no. 7839, pp. 604–609, 2020.

[4] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2023, pp. 1–22.

[5] Y. Duan and M. J. Wainwright, "Taming" data-hungry" reinforcement learning? stability in continuous state-action spaces," *arXiv preprint arXiv:2401.05233*, 2024.

[6] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.

[7] R. Chrisley, "Embodied artificial intelligence," *Artificial intelligence*, vol. 149, no. 1, pp. 131–150, 2003.

[8] Y. LeCun, "A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27," *Open Review*, vol. 62, no. 1, 2022.

[9] N. Zubić, F. Soldá, A. Sulser, and D. Scaramuzza, "Limits of deep learning: Sequence modeling through the lens of complexity theory," *arXiv preprint arXiv:2405.16674*, 2024.

[10] D. Hafner, J. Pasukonis, J. Ba, and T. Lillicrap, "Mastering diverse domains through world models," *arXiv preprint arXiv:2301.04104*, 2023.

[11] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review*, vol. 65, no. 6, p. 386, 1958.

[12] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[13] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.

[14] O. E. L. Team, A. Stooke, A. Mahajan, C. Barros, C. Deck, J. Bauer, J. Sygnowski, M. Trebacz, M. Jaderberg, M. Mathieu *et al.*, "Open-ended learning leads to generally capable agents," *arXiv preprint arXiv:2107.12808*, 2021.

[15] A. Latyshev and A. I. Panov, "Intrinsic motivation in model-based reinforcement learning: A brief review," *arXiv preprint arXiv:2301.10067*, 2023.

[16] B. Takkouche and G. Norman, "Prisma statement," *Epidemiology*, vol. 22, no. 1, p. 128, 2011.

[17] S. Chen, P.-L. Guhur, C. Schmid, and I. Laptev, "History aware multimodal transformer for vision-and-language navigation," *Advances in neural information processing systems*, vol. 34, pp. 5834–5847, 2021.

[18] H. Du, X. Yu, and L. Zheng, "Vtnet: Visual transformer network for object goal navigation," *arXiv preprint arXiv:2105.09447*, 2021.

[19] T. Chen, S. Gupta, and A. Gupta, "Learning exploration policies for navigation," *arXiv preprint arXiv:1903.01959*, 2019.

[20] A. Das, S. Datta, G. Gkioxari, S. Lee, D. Parikh, and D. Batra, "Embodied question answering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1–10.

[21] J. Lindblom, "A radical reassessment of the body in social cognition," *Frontiers in Psychology*, vol. 11, p. 484818, 2020.

[22] L. Soros and K. Stanley, "Identifying necessary conditions for open-ended evolution through the artificial life world of chromaria," in *Artificial Life Conference Proceedings*. MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info ..., 2014, pp. 793–800.

[23] O. Sigaud, G. Baldassarre, C. Colas, S. Doncieux, R. Duro, N. Perrin-Gilbert, and V.-G. Santucci, "A definition of open-ended learning problems for goal-conditioned agents," *arXiv preprint arXiv:2311.00344*, 2023.

[24] D. W. McShea, "Perspective metazoan complexity and evolution: is there a trend?" *Evolution*, vol. 50, no. 2, pp. 477–492, 1996.

[25] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

[26] E. Hughes, M. Dennis, J. Parker-Holder, F. Behbahani, A. Mavalankar, Y. Shi, T. Schaul, and T. Rocktaschel, "Open-endedness is essential for artificial superhuman intelligence," *arXiv preprint arXiv:2406.04268*, 2024.

[27] B. F. Skinner, *The behavior of organisms: An experimental analysis*. BF Skinner Foundation, 2019.

[28] C. L. Hull, "Principles of behavior: an introduction to behavior theory." *APA PsycInfo*, 1943.

[29] H. F. Harlow, "Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys." *Journal of comparative and physiological psychology*, vol. 43, no. 4, p. 289, 1950.

[30] P. Redgrave and K. Gurney, "The short-latency dopamine signal: a role in discovering novel actions?" *Nature reviews neuroscience*, vol. 7, no. 12, pp. 967–975, 2006.

[31] A. Saravanan, M. Tsfasman, M. A. Neerincx, and C. Oertel, "Giving social robots a conversational memory for motivational experience sharing," in *2022 31st IEEE international conference on robot and human interactive communication (RO-MAN)*. IEEE, 2022, pp. 985–992.

[32] T. Kulak and S. Calinon, "Combining social and intrinsically motivated learning for multitask robot skill acquisition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 15, no. 2, pp. 385–394, 2021.

[33] X. Lan, Y. Qiao, and B. Lee, "Multiagent hierarchical reinforcement learning with asynchronous termination applied to robotic pick and place," *IEEE Access*, 2024.

[34] M. Hasan and R. Niyogi, "Reward specifications in collaborative multi-agent learning: A comparative study," in *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, 2024, pp. 1007–1013.

[35] A. Bhoopchand, B. Brownfield, A. Collister, A. Dal Lago, A. Edwards, R. Everett, A. Fréchette, Y. G. Oliveira, E. Hughes, K. W. Mathewson *et al.*, "Learning few-shot imitation as cultural transmission," *Nature Communications*, vol. 14, no. 1, p. 7536, 2023.

[36] L. McNally, S. P. Brown, and A. L. Jackson, "Cooperation and the evolution of intelligence," *Proceedings of the Royal Society B: Biological Sciences*, vol. 279, no. 1740, pp. 3027–3034, 2012.

[37] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[38] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton *et al.*, "Mastering the game of go without human knowledge," *nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[39] M. F. A. R. D. T. (FAIR)†, A. Bakhtin, N. Brown, E. Dinan, G. Farina, C. Flaherty, D. Fried, A. Goff, J. Gray, H. Hu *et al.*, "Human-level play in the game of diplomacy by combining language models with strategic reasoning," *Science*, vol. 378, no. 6624, pp. 1067–1074, 2022.

[40] T. Bartz-Beielstein, J. Branke, J. Mehnen, and O. Mersmann, "Evolutionary algorithms," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 4, no. 3, pp. 178–195, 2014.

[41] Y. L. B. Nogueira, C. E. F. de Brito, C. A. Vidal, and J. B. Cavalcante-Neto, "Towards intrinsic autonomy through evolutionary computation," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4449–4473, 2020.

[42] J. C. Brant and K. O. Stanley, "Minimal criterion coevolution: a new approach to open-ended search," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2017, pp. 67–74.

[43] J. Gomes, P. Mariano, and A. L. Christensen, "Novelty-driven cooperative coevolution," *Evolutionary computation*, vol. 25, no. 2, pp. 275–307, 2017.

[44] A. Romero, F. Bellas, and R. J. Duro, "An autonomous drive balancing strategy for the design of purpose in open-ended learning robots." in *AAMAS*, 2021, pp. 1631–1633.

[45] G. Paolo, A. Coninx, S. Doncieux, and A. Laflaquière, "Sparse reward exploration via novelty search and emitters," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2021, pp. 154–162.

[46] J. K. Pugh, L. B. Soros, and K. O. Stanley, "Quality diversity: A new frontier for evolutionary computation," *Frontiers in Robotics and AI*, vol. 3, p. 202845, 2016.

[47] R. Portelas, C. Colas, L. Weng, K. Hofmann, and P.-Y. Oudeyer, "Automatic curriculum learning for deep rl: A short survey," *arXiv preprint arXiv:2003.04664*, 2020.

[48] S. Forestier, R. Portelas, Y. Mollard, and P.-Y. Oudeyer, "Intrinsically motivated goal exploration processes with automatic curriculum learning," *Journal of Machine Learning Research*, vol. 23, no. 152, pp. 1–41, 2022.

[49] P. Dhakan, K. Kasmarik, I. Rañó, and N. Siddique, "Open-ended continuous learning of compound goals," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 2, pp. 274–285, 2019.

[50] G. Wang, Y. Xie, Y. Jiang, A. Mandlekar, C. Xiao, Y. Zhu, L. Fan, and A. Anandkumar, "Voyager: An open-ended embodied agent with large language models," *arXiv preprint arXiv:2305.16291*, 2023.

[51] M. Reynard, "Reinforcement learning in the minecraft gaming environment," Ph.D. dissertation, Stellenbosch: Stellenbosch University, 2020.

[52] A. Barto, M. Mirolli, and G. Baldassarre, "Novelty or surprise?" *Frontiers in psychology*, vol. 4, p. 61898, 2013.

[53] G. Baldassarre, "Intrinsic motivations and open-ended learning," *arXiv preprint arXiv:1912.13263*, 2019.

[54] Y. Gatsoulis, C. Burbridge, and T. M. McGinnity, "Biologically inspired intrinsically motivated learning for service robots based on novelty detection and habituation," in *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2012, pp. 464–469.

[55] X. Cao, T. Lu, and Y. Cai, "Intrinsic motivation for deep deterministic policy gradient in multi-agent environments," in *2020 Chinese Automation Congress (CAC)*. IEEE, 2020, pp. 1628–1633.

[56] O. Loyola, J. Kern, and C. Urrea, "Novel algorithm for agent navigation based on intrinsic motivation due to boredom," *Information Technology and Control*, vol. 50, no. 3, pp. 485–494, 2021.

[57] V. R. Kompella, M. Stollenga, M. Luciw, and J. Schmidhuber, "Continual curiosity-driven skill acquisition from high-dimensional video inputs for humanoid robots," *Artificial Intelligence*, vol. 247, pp. 313–335, 2017.

[58] J. Schmidhuber, "Curious model-building control systems," in *Proc. international joint conference on neural networks*, 1991, pp. 1458–1463.

[59] N. C. Volpi and D. Polani, "Goal-directed empowerment: combining intrinsic motivation and task-oriented behavior," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 15, no. 2, pp. 361–372, 2020.

[60] V. G. Santucci, G. Baldassarre, and M. Mirolli, "Grail: A goal-discovering robotic architecture for intrinsically-motivated learning," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 3, pp. 214–231, 2016.

[61] V. G. Santucci, D. Montella, and G. Baldassarre, "C-grail: Autonomous reinforcement learning of multiple and context-dependent goals," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 15, no. 1, pp. 210–222, 2022.

[62] O. L. Georgeon, J. B. Marshall, and S. Gay, "Interactional motivation in artificial systems: Between extrinsic and intrinsic motivation," in *2012 IEEE international conference on development and learning and epigenetic robotics (ICDL)*. IEEE, 2012, pp. 1–2.

[63] R. Agarwal, M. Schwarzer, P. S. Castro, A. C. Courville, and M. Bellemare, "Deep reinforcement learning at the edge of the statistical precipice," *Advances in neural information processing systems*, vol. 34, pp. 29 304–29 320, 2021.

[64] N. Duminy, D. Duhaut *et al.*, "Strategic and interactive learning of a hierarchical set of tasks by the poppy humanoid robot," in *2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. IEEE, 2016, pp. 204–209.

[65] C. Gan, A. Bhandwaldar, A. Torralba, J. B. Tenenbaum, and P. Isola, "Open: An open-ended physics environment for learning without a task," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 5878–5885.

[66] G. Baldassarre, D. Montella, V. G. Santucci, and E. Cartoni, "Real 2021–robot open-ended autonomous learning: A competition and benchmark," in *2021 IEEE International Conference on Development and Learning (ICDL)*. IEEE, 2021, pp. 1–8.

[67] B. Fang, Z. Peng, H. Sun, and Q. Zhang, "Meta proximal policy optimization for cooperative multi-agent continuous control," in *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2022, pp. 1–8.

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

[69] A. Kanervisto, S. Milani, K. Ramanauskas, N. Topin, Z. Lin, J. Li, J. Shi, D. Ye, Q. Fu, W. Yang *et al.*, "Minerl diamond 2021 competition: Overview, results, and lessons learned," *NeurIPS 2021 Competitions and Demonstrations Track*, pp. 13–28, 2022.

[70] J. C. Brant and K. O. Stanley, "Benchmarking open-endedness in minimal criterion coevolution," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2019, pp. 72–80.

[71] O. Mees, L. Hermann, E. Rosete-Beas, and W. Burgard, "Calvin: A benchmark for language-conditioned policy learning for long-horizon robot manipulation tasks," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7327–7334, 2022.

[72] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[73] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Doll'a r, and C. L. Zitnick, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014. [Online]. Available: http://arxiv.org/abs/1405.0312

[74] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[75] M. Abdin, S. A. Jacobs, A. A. Awan, J. Aneja, A. Awadallah, H. Awadalla, N. Bach, A. Bahree, A. Bakhtiari, H. Behl *et al.*, "Phi-3 technical report: A highly capable language model locally on your phone," *arXiv preprint arXiv:2404.14219*, 2024.

[76] Z. Ge, H. Huang, M. Zhou, J. Li, G. Wang, S. Tang, and Y. Zhuang, "Worldgpt: Empowering llm as multimodal world model," *arXiv preprint arXiv:2404.18202*, 2024.

[77] J. Bruce, M. Dennis, A. Edwards, J. Parker-Holder, Y. Shi, E. Hughes, M. Lai, A. Mavalankar, R. Steigerwald, C. Apps *et al.*, "Genie: Generative interactive environments," *arXiv preprint arXiv:2402.15391*, 2024.

[78] C. Wirth, R. Akrour, G. Neumann, and J. Fürnkranz, "A survey of preference-based reinforcement learning methods," *Journal of Machine Learning Research*, vol. 18, no. 136, pp. 1–46, 2017.

[79] Z. Guo, M. Yao, Y. Yu, and Q. Yin, "Improve the efficiency of deep reinforcement learning through semantic exploration guided by natural language," *arXiv preprint arXiv:2309.11753*, 2023.

[80] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski, T. Ding, D. Driess, A. Dubey, C. Finn *et al.*, "Rt-2: Vision-language-action models transfer web knowledge to robotic control," *arXiv preprint arXiv:2307.15818*, 2023.

[81] J.-A. Meyer and S. W. Wilson, *A Possibility for Implementing Curiosity and Boredom in Model-Building Neural Controllers*. IEEE, 1991, pp. 222–227.

[82] N. Chentanez, A. Barto, and S. Singh, "Intrinsically motivated reinforcement learning," *Advances in neural information processing systems*, vol. 17, 2004.

[83] D. Hafner, "Benchmarking the spectrum of agent capabilities," *arXiv preprint arXiv:2109.06780*, 2021.

A Google Scholar articles

Table 1: Grouping of Google Scholar papers based on primary focus.

Category	Source
Intrinsic motivation	[58] [52] [53] [19] [81] [82]
Open-ended learning	[50] [47] [26] [27] [14] [18] [23]
Benchmarks and datasets	[69] [68] [83] [71] [63]

B Usage of Large Language Models

In the context of this literature review, GPT-3.5[8] has been used to analyze and summarize different scientific articles. No Language Model has been used for writing the manuscript.

[8]https://chatgpt.com/