Affective Reinforcement Learning: A Taxonomy and Survey

Matthew Barthet *IEEE Member*, Ahmed Khalifa, Antonios Liapis, and Georgios N. Yannakakis, *IEEE Fellow*
Institute of Digital Games, University of Malta
Msida, Malta
matthew.barthet@um.edu.mt, ahmed.khalifa@um.edu.mt, antonios.liapis@um.edu.mt,
georgios.yannakakis@um.edu.mt

***Abstract*—This paper surveys the current state-of-the-art of reinforcement learning methods and principles applied to affective computing across a wide variety of domains. We review this nascent field, which we refer to as *affective reinforcement learning*, that interweaves core RL components within the affective loop. Specifically, we survey the use of affective information across the three major components of the reinforcement learning (RL) paradigm: a) shaping a *reward* signal, b) driving the *action* policy, and c) forming part of the *state* representation. We introduce a taxonomy of different terms for affective RL, and we conclude by discussing the current limitations of the framework as well as several open and promising research directions within this emerging area.**

***Index Terms*—affective computing, reinforcement learning, survey, taxonomy**

I. Introduction

The challenging nature of accurately modeling subjective notions such as human affect remains a significant hurdle to overcome towards realizing human-centered artificial intelligence (AI). However, as AI methods—such as the transformer architecture [1]—become more effective in recognizing complex spatiotemporal phenomena, affective computing (AC) [2] systems become increasingly deployable across various domains, including healthcare [3], vehicular systems [4], and entertainment [5]. Such affective interactions are normally represented as an *affective loop* [6] process by which a human user is exposed to stimuli that elicit emotional responses which can be detected through various forms of verbal or non-verbal cues. The predicted emotions can then be used to adjust the interaction by presenting a new set of stimuli to the user. Within the affective loop, the process of detecting and modeling affect is predominantly viewed from a supervised learning (SL) lens [5], [7], [8]. However, deploying such models to real-world environments typically requires the model to act, adapt, and tailor itself to a user's affective patterns, which is a challenge in its own right, particularly for a static SL model. Viewing affective interaction from a reinforcement learning (RL) [9] lens offers several benefits not only for representing and understanding human affect but also for the successful deployment of affective interaction in the wild.

In this paper, we introduce and survey the emerging research area that interweaves aspects of affective interaction within the RL paradigm, which we name *affective reinforcement learning* (ARL) framework; see Fig. 1. The framework is advantageous for both AC and RL research for a number of reasons: First, the RL paradigm is heavily inspired by biological processes and psychological phenomena in a way that aligns closely with the ways humans learn and affectively interact with their surroundings [9]. This makes RL particularly well-suited for learning affective interactions relying on theoretical models of emotion, which are predominantly used in AC such as the OCC model [10]. Second, RL tends to tackle domains that are highly relevant to AC research, such as social robotics [11], making ARL a natural overlap between the field (AC) and the method (RL). Finally, RL stands to benefit from AC research since recent evidence suggests that emotion can be a powerful signal for both learning transferable behaviors and improving the transparency of the agent's actions [12], [13].

There are only a handful of papers surveying the intersection of AC with RL [3], [14], [15] mostly focusing, however, on a particular domain or AC use case. This includes survey papers on affect models for video games [5]—briefly covering the niche use of training RL agents—papers that focus on RL for training virtual agents and robots using *artificial* emotions [14], thorough studies that investigate *empathy* in virtual agents [15], survey papers on the use of RL within the domain of healthcare [3] and, finally, papers with a focus on RL in conjunction with various forms of human feedback [16] but limited emphasis of affective signals. In this paper, we attempt to fill the aforementioned gap in the literature by offering a comprehensive, up-to-date, and holistic overview of the state of the art in the intersection of RL and affective computing; moreover, we introduce a taxonomy for this emerging area. We put emphasis on how the various examples in academic literature and industrial practice build on human and artificial emotion representations with respect to both the *affective loop* [6] and the Markov Decision Process (MDP) [9]. We also offer a high-level overview of the most popular domains in which ARL is currently being applied, and we outline how the area can take the next steps towards safer, more believable, and emotionally aware virtual agents and generative AI processes.

The remainder of the paper is structured as follows. First we define our ARL framework and give an overview of the intersection between RL and the affective loop in Section II. Our survey first investigates the various ways affect has been used to build reward functions (see Section III) with an

Manuscript received April 19, 2005; revised August 26, 2015.

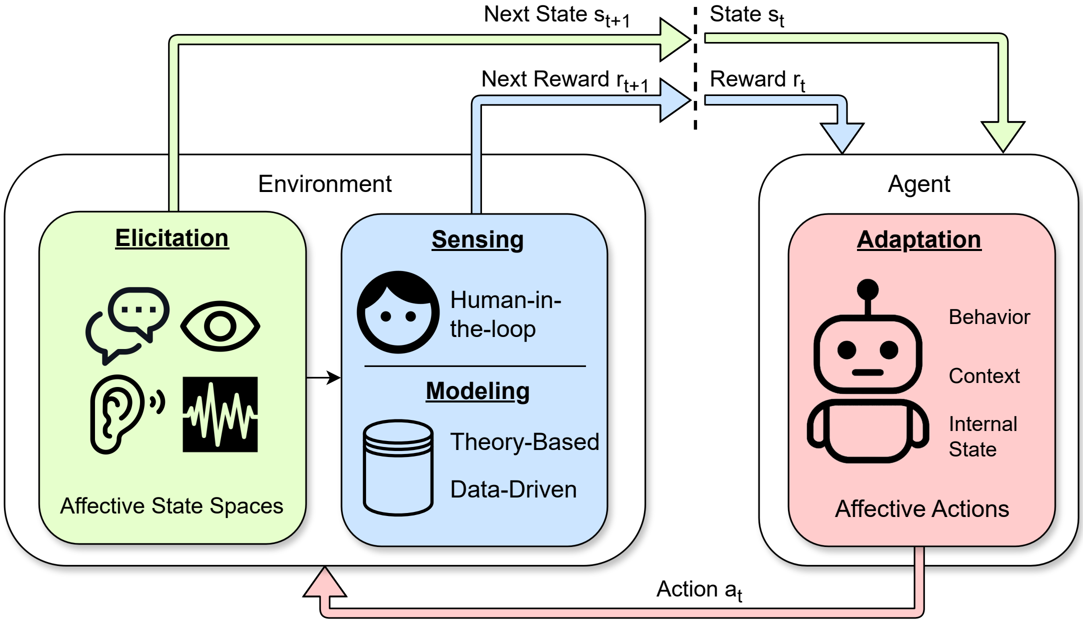


Fig. 1. The affective loop framed within the reinforcement learning framework and the RL elements that are directly used by the affective loop. Green elements correspond to the observation data used by the agents to learn, blue elements correspond to the rewards derived from human affect data (either via a human-in-the-loop or a model-based approach), or feedback from the environment itself (e.g. increase in score in a game), and red elements correspond to the potential action mechanisms of the agents. The elements of the affective loop are indicated using bold and underlined text.

emphasis on the different processes for capturing affect as a reward signal. We then survey the different types of action types learned by affective RL agents (see Section IV). In Section V, we categorize approaches based on the observation spaces used and highlight studies that incorporate affect as part of the state representation. Finally, in Section VI we identify a number of open directions that we view as the most promising for further research in this emerging area.

II. Affective RL: A Unified Approach

The introduced affective RL framework integrates the phases of the affective loop [6] into the RL paradigm and demonstrates how affect can be incorporated to train RL agents, as illustrated in Fig. 1. Viewing the affective loop under an RL lens offers a number of important benefits for the study of affect. First, ARL can be used to train more human-like and emotion-aware agents, or use affect as a reward signal to infer more generalizable policies that would be otherwise difficult to infer using human-authored reward functions. Second, ARL can offer a more data-efficient approach compared to the dominant supervised learning paradigm in affect modeling. The framework does not require large affective corpora; a recent study found that such datasets can include anywhere from 3 to 250 human participants when used to train deep machine-learning models [17]. Finally, ARL can leverage multiple training signals, including human demonstrations, reward signals derived from theoretical models on psychology and affective sciences, pre-trained models of affect, or even a combination of such signals.

Figure 1 illustrates how the ARL framework builds upon and interweaves the four phases of the affective loop (bold text) and the RL paradigm (colored elements). The remainder of this section is organized as follows. First, we present the Markov Decision Process (MDP) and the dominant definition of RL, which we use in ARL (see Section II-A). Then, in Section II-B, we define the affective loop and tie in how ARL utilizes different phases of the loop, which, in turn, forms the structure of the survey. We end the section (see Section II-C) with an outline of the methodological approach we took for completing this survey.

A. Reinforcement Learning as Affective Interaction

Reinforcement learning is a well-established paradigm of machine learning inspired by the way biological learning occurs through rewards and penalties [9]. The RL problem is framed as a closed-loop system in which an agent continually interacts with an environment to learn to achieve a specific goal. More specifically, the agent repeatedly selects an *action* ($a_t \in A$), at discrete time steps (t) according to its current *state* ($s_t \in S$)—as provided by the environment—in order to maximize an expected accumulated *reward* (R_t) signal (see Fig. 1). The environment then transitions to a new state (s_{t+1}) according to a transition function, and provides the agent an immediate reward (r_{t+1}) that quantifies the desirability of the chosen action in that state. The agent's objective is to learn a policy ($\pi(a_t|s_t)$)—or a mapping from any given state to any possible action—that maximizes its expected accumulated reward over time. We use the three major components of the RL paradigm—state, action, and reward—as core dimensions of the ARL studies we surveyed, detailed further below:

- **State:** In ARL, the states provided by the environment contain affective information embedded within their features, which the agent uses to learn an optimal, affective policy. We distinguish between three types of state spaces used to train affective agents (see Section V). These are contextual features specific to the environment, and verbal or non-verbal cues provided by humans or other agents.

- **Action:** Affective agents typically fall into three main categories according to the types of actions they take, described in Section IV. The most common form of action governs agent *behavior* within an environment—such as playing a game, interacting with a patient, or driving a vehicle. Agents can also create their *context* by generating entirely new interactions and experiences for other agents or humans in the environment, such as new game levels [18] or new stimuli for exposure therapy [19]. Finally, agents can take actions that change their *internal* affective state rather than take external actions described earlier.
- **Reward:** The reward signal is the most specific component of the RL framework, as it defines the intended behavior of the environment and relies heavily on the agent's task and the nature of the interaction. In ARL, the reward typically directly contains affect data obtained via two approaches: a *human-in-the-loop* approach through implicit sensing (see Section III-A) or explicit sensing (see Section III-B), or a *model-based* approach using data-driven models (see Section III-C) or theoretical models (see Section III-D).

RL differs from supervised learning—the dominant affect modeling paradigm—in that it does not rely on labeled input–output pairs. Instead, the agent must explore the environment to discover which actions yield higher rewards, making exploration–exploitation trade-offs a central challenge. This distinction makes RL particularly suitable for sequential decision-making problems, including robotics, healthcare, autonomous systems, and games. The formulation of the reward function is central to the success of an RL agent. Poorly designed reward functions can lead to suboptimal or unintended behavior, especially in complex environments. This has led to growing interest in incorporating richer forms of feedback—including temporal affective signals—as a way to shape agent behavior in alignment with human preferences, values, and emotional responses [20]. Such approaches extend the traditional RL framework by embedding affect as an intrinsic or extrinsic motivation signal, placed at the core of the ARL framework. As a result, ARL is capable of training agents that are more capable of capturing and manifesting human-like behavior and emotion, whilst also being potentially more generalizable across tasks and environments [20].

The RL algorithms we chose to include in ARL can be broadly categorized into value-based, policy-based, and actor-critic methods. Value-based methods, such as Q-learning [21], estimate the return of taking a given action in a given state, and derive the policy indirectly by acting greedily with respect to these value estimates. Policy-based methods, in contrast, directly optimize the policy using gradient-based methods such as REINFORCE [22]. Actor-critic approaches, such as Asynchronous Advantage Actor Critic (A3C) [23] or Proximal Policy Optimization (PPO) [24], are the most versatile and dominant method [25], [26] by combining both paradigms, using a value function—referred to as the *critic*—that guides policy updates performed by the *actor*, often leading to more stable and sample-efficient learning. We also include algorithms that operate at the fringes of the traditional RL defined above. Such examples include exploration-based algorithms such as Go-Explore [27], curiosity-driven learning [28], and unsupervised skill discovery [29]. While these approaches may not fit precisely under the standard RL definition, they are often used as a precursor or augmentation to traditional RL training and are highly relevant in the context of affective learning—particularly in sparse or ambiguous reward settings.

B. The Affective Loop as an RL Process

The *affective loop* [6] is a well-established paradigm that is used to describe affective interactions in a generalized, high-level manner. The loop can be broken down into a sequence of four phases, as shown by the bold and underlined text in Fig. 1. To design our ARL framework, we build upon the reinforcement learning loop and seek possible implementations of its different components (state, action, and reward) using the affection loop phases. This makes the affective loop phases as a secondary dimension for our ARL framework main phases (as shown in Fig. 1. These four phases are as follows:

1) **Elicitation:** The first phase of the loop is where the environment provides a stimulus to the agent and any human in-the-loop, if present. The type of stimuli could take many forms depending on the domain being tackled. The agent collects observations based on these stimuli across one or more modalities, which we describe in Section V. The stimuli presented to the agent are the driver of the agent's learning in terms of its action and its affective response as generated later on in the loop.
2) **Sensing:** Affect can be sensed in real-time through a human in the loop. This can either be provided through implicit feedback (see Section III-A) or explicit feedback (see Section III-B).
3) **Modeling:** If it is not plausible to integrate a human-in-the-loop, a *model-based* approach can be used to predict affect based on the observations collected by the agent. Such models can be either *data-driven* using machine learning or large pre-trained models (see Section III-C) or *theory-based*, which rely on ad-hoc designed reward functions for approximating affect (see Section III-D).
4) **Adaptation:** In the final phase of the loop, the agent reacts to the outcome of the previous timestep, and uses the reward provided by the environment to modify its internal policy to maximize its expected return over time. To close the loop, the agent passes its selected action to the environment, which, in turn, executes the next timestep and provides a new set of stimuli to the agent.

RL presents several benefits to the realization of the affective loop and the development of reliable and transparent affective systems. The RL paradigm leans heavily on biology and psychology, making it well aligned with how humans perceive and interact with their environment. We believe this makes it naturally well-suited for training agents to learn affective behaviors and facilitate richer human-computer interaction.

Furthermore, many of the domains actively researched within RL are highly relevant to affective computing research, such as healthcare [3], robotics [30], and games [5]. Given the overlap and natural fit between these two research fields, we believe ARL defines a promising new direction for affective computing research and human-computer interaction.

C. *Affective RL Survey*

Our survey includes 50 papers that fall under the ARL framework and are discussed in the remainder of this paper. We identified these papers as follows. First, we used Arxiv's API to retrieve any papers that included the terms "reinforcement learning" and "affective computing", or "emotion" or "player experience" or "procedural content generation" within their title or abstract. We include *player experience* and *user experience* as they are often used to refer to the emotional experiences of users within games research and human-computer interaction, respectively. We also included *procedural content generation* to locate potential papers that use generative AI methods in association with RL and affect. This initial process yielded 105 papers for further review. We then performed an additional search using Google Scholar using the same keywords to pick up on any literature not hosted on Arxiv, and other works identified within the reference list of existing papers. Finally, we removed any vision papers as well as any papers that were found to be off topic to reach the final count of 50 papers.

III. Affective Reward

The ARL loop (see Fig. 1) differs from traditional reinforcement learning mainly due to the way that *rewards to the agent capture human affect*. We distinguish between rewards sensed from *humans* involved during the agent training process (i.e. observing the environment and state changes), and rewards based on *models* trained from human data (collected a priori) or derived from theory. For the former case, we borrow the term *human-in-the-loop* [31], [32] from interactive machine learning to describe it, while for the latter case, we call it Model-based. For the human-in-the-loop, we differentiate between two cases: Implicit and explicit. *Implicit* feedback contains relevant affect information as a side effect of actions typically provided for other purposes [16], rather than *explicit* feedback where the sole purpose of a person's action is to provide feedback. For the Model-based, we differentiate between two cases: models based on human data that were collected a priori (called Data-Driven Modeling) or models that are based on theory derived from psychology (called Theory-driven Modeling).

A. *Human-in-the-Loop: Implicit Sensing*

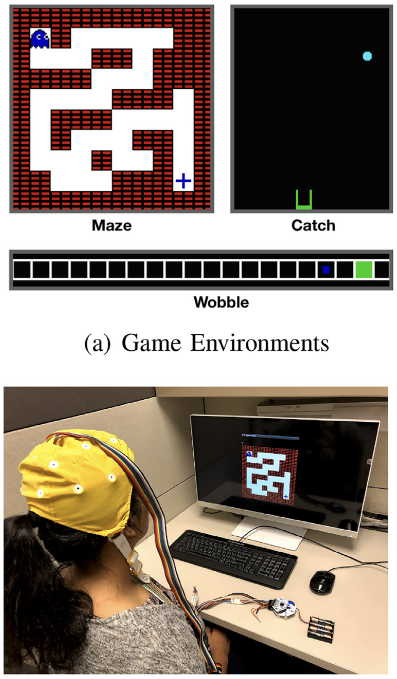
The majority of RL studies using *implicit* feedback capture affect via specialized sensors that record users' physiological signals [33]. We will use the signals themselves to group the ARL work in this vein, which fall into brain activity via *electroencephalogram* and skin conductance via *electrodermal activity*; we finish with studies using non-physiological feedback.

The *electroencephalogram* (EEG) signal measures brain activity in a non-invasive manner by sensing electrical activity in the scalp, and is commonly used in brain-computer interfaces (BCI) [34] for domains such as robotics [35] and medicine [36]. In ARL, real-time raw EEG signals have been used to monitor drowsiness in road safety applications. Ming et al. [37] processed EEG signals to extract features using a deep Q-network to predict whether reaction time will increase or decrease in the current time window. The agent is rewarded based on the negative absolute error between the predicted and actual reaction times for each event. In a similar application, Yousaf et al. [38] trained a deep RL agent using EEG features captured from edge devices to classify the driver's cognitive state as attentive, inattentive, or drowsy. The agent could play auditory alerts or adjust vehicle speed to sustain or improve driver attention, receiving penalties if the driver's focus deteriorates.

A common way to process EEG signals is through error-related potentials (ErrP), which are naturally occurring event-driven signals in the brain that fire after an error is perceived [39]. ErrP can act as an efficient proxy for a reward function to penalize an agent's erroneous behaviors. Prior to RL, ErrP signals were applied in simple state-action spaces such as a cursor control task [33] or a binary decision-making task in robotics [40]. Learning happens by having a human participant observe the agents behave whilst wearing an EEG cap, which feeds raw signals into an ErrP classifier, which flags when the user detects incorrect behavior. A positive ErrP (1) indicates the agent should change its behavior, whereas an absence of ErrP (0) indicates the agent is performing well. Akinola et al. [41] leveraged ErrP signals for sparse reward environments in a vehicle navigation task, outperforming conventional sparse-reward RL. Kim et al. [42] trained a robot arm to pick up and place objects on a table in MuJoCo, and found that agents trained via ErrP outperform sparse-reward RL agents and compare favorably to fully engineered densely rewarded agents. These results highlight that ErrP could remove the burden of designing complex reward functions by leveraging real-time human sensing alone.

A notable study in ErrP signals for RL rewards used multiple grid-based game environments to show the generalization potential of these signals [20]. Xu et al. [20] invited participants to mentally assess the performance of RL agents in three games while wearing an electrode cap (see Fig. 2). In the first game, the agent must learn to move itself (a cursor) along one dimension toward a target cell, as seen in previous studies [33]. In the second game, the agent moves itself (a bucket) along one dimension and must catch a falling ball. In the third game, the agent navigates a two-dimensional maze to reach the target. They perform a zero-shot learning study across the three games, illustrating the universality of ErrP signals across environments. Their results show that giving agents full access to ErrP signals in the training pipeline speeds up learning, and that ErrP signals detected using a classifier trained on one game can be used to accurately train an agent in other games.

Skin conductance is commonly captured through *electrodermal activity* (EDA), and has shown strong predictive power



(a) Game Environments

(b) Experiment Bench

Fig. 2. Implicit feedback for training gameplaying agents for a maze navigation and a ball catching game (top). Feedback is provided through ErrP signals derived from EEG signals using a human-in-the-loop affect reward scheme (bottom). Image taken from [20] with authors' permission.

for emotions such as surprise and fear [43]. Real-time EDA signals have been used to train agents to generate personalized exposure therapy stimuli for arachnophobia [19]. In that work, EDA served as a proxy for anxiety, and was used as a reward for a designer agent that adjusted attributes of a virtual spider (e.g. size, color, locomotion). The designer agent's goal was to elicit stronger anxiety (manifested through EDA responses) in patients.

Beyond physiological signals, unobtrusive signals such as *speech* can also implicitly capture human affect. Kim and Scassellati [44] relied on prosody (i.e. the rhythm, sound, and intonation of language) to infer affect from a human teacher in order to train a robot to wave its arm in a desired manner. Weber et al. [45] trained a robot to adapt its jokes to participants' sense of humor by screening for laughter and changes in facial expression. Similarly, [46] uses facial emotion recognition from humans-in-the-loop to train a drone to fly by rewarding the agent for eliciting positive expressions.

B. Human-in-the-Loop: Explicit Sensing

Explicit feedback involves the participant providing intentional and direct guidance to the agent. For affect, this usually falls under a self-reporting interface using ratings or labels [47]. This is the most direct and simple type of affect labeling, but it comes at the cost of great cognitive demand on the part of the human annotator. Reaction times and noise can also become a significant issue, especially with challenging internal manifestations such as emotions. In practice, we have found only tangential work that uses explicit human feedback for ARL.

A notable, if tangential, example of explicit rewards is presented by Shaik et al. [48]. They propose a multi-agent system for patient monitoring: each agent monitors a different signal (i.e. heart rate, respiration, temperature) as part of the agent's state space. The actions the system takes may alert the appropriate medical emergency team based on escalations in pre-defined health parameters. The system is trained via Deep Q-Networks based on errors from the environment (i.e. wrong or correct alerts). Affective interaction is designed for human-in-the-loop rewards, with the medical emergency teams providing the rewards for correct or incorrect alerts; however, the proposed system was evaluated solely on existing datasets with real-world physiological and motion data using ad-hoc thresholds for emergency responses. It is important to note that even in ideal situations, the human-in-the-loop protocol would not provide *affect* rewards per se, as whether the alert was correct or not is a purely cognitive task. However, this study is the closest example of a system we were able to identify that uses both sensors for monitoring the environment and explicit (if not affect-based) rewards to guide the agent.

C. Data-driven Modeling

As discussed in Section III-A, relying on explicit human feedback during training is a very cumbersome process. Instead, most ARL approaches use human data collected offline (prior to the agent training task) and build affect models that can be used as rewards during training. These models do not have to use state-of-the-art AI: we group approaches for data-driven affect rewards based on the machine learning algorithm used. This leaves us with three groups: *proximity-based*, *traditional machine learning*, and *transformer-based* approaches.

Depending on the volume and format of available data, some affect corpora can be used without elaborate machine learning model training. In their *Go-Blend* [49] algorithm, Barthet et al. used a k-Nearest Neighbors model to approximate a player's arousal. Their study trained an agent to play a 3D driving game, based on rewards of arousal (paired or not with in-game score rewards). Mapping the agent's current game state to the nearest game state of human playthroughs was possible due to a large corpus of time-continuous arousal annotations paired with real-time game metrics available from the AGAIN dataset [50]. Agent rewards were calculated as the similarity between the agent's arousal and a target arousal signal based on player personas [51], [52]. Barthet et al. [53] experimented with this method further by rewarding the agent for visiting high arousal states in an attempt to learn more optimal behavior across three games: a driving game, a platformer game, and a first-person shooter. An illustration of this data-driven approach is depicted in Fig. 3.

Traditional machine learning approaches, mostly variants of recurrent neural networks (RNNs), have often been used to build data-driven models for affect rewards. EmoRL [54] trains an agent to decide when to make predictions of anger in human participants using an RNN trained on the IEMOCAP

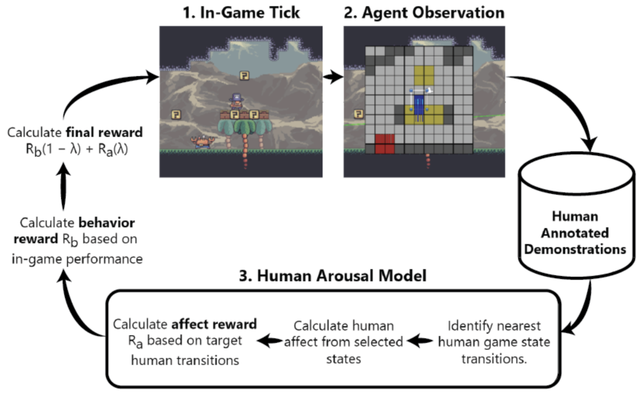

1. In-Game Tick
2. Agent Observation
Human Annotated Demonstrations
3. Human Arousal Model
Identify nearest human game state transitions.
Calculate human affect from selected states
Calculate **affect reward** R_a based on target human transitions
Calculate **behavior reward** R_b based on in-game performance
Calculate **final reward** $R_b(1 - \lambda) + R_a(\lambda)$

Fig. 3. Example of a data-driven approach to affective rewards. Human affect demonstrations can be used to create an affect model which generates a reward signal. The reward can be used to train the agent, or combined with other signals such as behavioral rewards.

corpus [55], rewarding for accuracy and latency. Li et al. [56] trained a Convolutional Neural Network (CNN) classifier to categorize replies into sentiment categories in order to train an RL agent to generate emotionally constrained text replies. The agent is rewarded for generating replies which are coherent, on topic, and emotionally relevant. Churamani et al. [57] trained an actor-critic RL agent to negotiate resources with a human player in the Ultimatum game, by using CNN models and self-organizing RNNs. The models were trained to predict arousal and valence values from audio-visual inputs of human participants, forming temporal representations that track user behavior over entire interactions. The agent is rewarded for improving its resources and causing a positive shift in its affective mood. Liu et al. [58] used a pre-trained speech emotion recognition model using a CNN Long-Short-Term Memory (LSTM) network as a reward signal to a policy gradient agent that creates speech expressing specific emotions.

The emergence of the transformer architecture and early pretrained large language models opened up new possibilities for ARL research. Unlike most previously mentioned approaches, which had to train the affect reward models on existing corpora, transformer-based reward models come pre-trained and thus could be used in a zero-shot manner or minimally refined through finetuning.

Early transformer architectures were small enough to fine-tune or train from scratch for tailored affect rewards. Shin et al. [59] employed pre-trained BERT classifiers, which they fine-tuned on datasets with emotionally labeled dialogue: SST-2 [60] and *EmpatheticDialogues* [61]. These emotion classifiers were used to train an agent to maximize empathy in its natural language responses via REINFORCE [22]. Jhain et al. [62] trained BERT models based on valence-arousal coordinates derived from the *EmpatheticDialogues* dataset, and used them to derive rewards to aplify empathy valence for deep RL agent training. Zhou et al. [63] trained conversational assistants via BERT, DistilBERT, ALBERT, and RoBERTa models [64] that classify user emotions into positive, negative, or neutral. Notably, Zhou et al. trained all models in an end-to-end fashion on a private dataset of emotionally labeled e-mails, rewarding the agent for eliciting positive emotions in simulated customers.

Brahman and Chaturvedi [66] fine-tuned a pre-trained GPT-2 model to reward a storytelling agent based on similarity to desired emotion arcs within the story. Sharma et al. [67] trained an agent to perform sentence-level edits in a story to increase empathy in posts and maintain text fluency, sentence coherence, context specificity, and diversity. While text quality rewards are computed through a combination of pre-trained and custom-trained language models, the affect reward (change in empathy) uses a custom-trained RoBERTa model on a dataset of paired interactions labeled for empathy [68]. Similarly, Ma et al. [69] use pretrained text-to-text transformers [70] to train an empathy identifier on a mental health corpus, which acts as a reward for a PPO agent trained to generate empathic responses. Finally, Rahman et al. [71] use the latent space of a graph-based transformer model as input to train a dueling DQN agent for emotion recognition, with a fixed positive or negative reward for correct or incorrect classifications.

While Large Language Models (LLMs) have been popular for a multitude of zero-shot tasks, their applications in ARL seem underexplored. The only example of truly large models for affect rewards is the work of Yuan et al. [72] for an agent assisting in dementia care. The agent detected the state of the patient—in terms of forgetfulness, confusion, anger, and disengagement—through a pre-trained LLM with no finetuning (GPT-4o), using the patient's states (along with other parameters such as timesteps or task completion) as a composite affect reward for the robotic assistant.

D. Theory-based Modeling

A common alternative to intrusive human feedback or data-hungry models for affect rewards is to manually define rules for affect based on theoretical frameworks of psychology and affective sciences, specifically. These reward signals are generally more interpretable and transparent than those derived from black-box models [73], whilst also being more lightweight computationally. However, their fixed and human-designed nature means they are generally less adaptable and highly specific to the context of the agent. This approach is common in the broader affective computing literature, such as matching laughter as an increase in joy [74] or encountering a monster in a horror game as an increase in tension [75].

An indicative example comes from bipedal robots handling the task of social navigation. Zhu et al. [76] rewarded the robot based on a variable minimum distance between the robot and pedestrians based on the pedestrians' emotions; both the values for comfortable social distance and how pedestrians' happy, neutral, or negative emotions affect such distance was based on psychological studies [77].

Within games and player modeling research [78], common affect constructs studies are those of fun, enjoyment, and engagement. Early studies on affective computing (beyond RL) used principles from game design [79] and the psychological concept of *flow* [80] to estimate enjoyment based on diversity

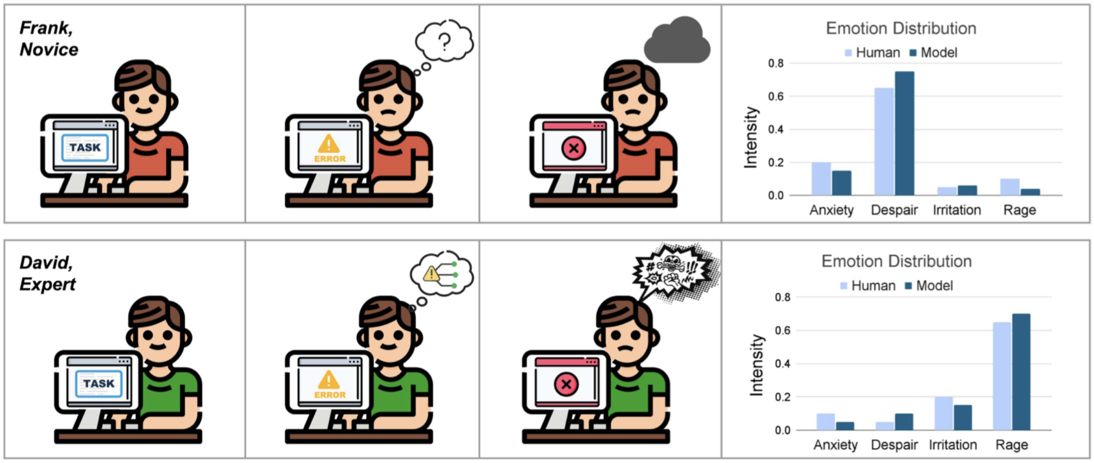


Fig. 4. Emotional responses to events depend on cognitive appraisals. For example, when faced with the same computer failure, Frank, a novice, feels desperation due to low perceived control, while David, an expert, appraises the situation with more power, though it still obstructs his goals. The model predicts emotions by evaluating factors like suddenness, goal relevance, conduciveness, and power, generating an appraisal vector that maps to emotions. Differences in responses mainly stem from perceived power, which in RL terms reflects an agent's ability to influence outcomes and rewards. Model predictions are matched with human data from vignette experiments. Image taken from [65] with authors permission.

of stimuli [81]. Shu et al. [18] leveraged similar notions for an agent that generates fun levels for *Super Mario Bros* (Nintendo, 1985). The agent's action was to generate the next segment of the level, and was rewarded based on a "sweet spot" of tile diversity with previous level segments (calculated via KL-divergence) as an estimation of Koster's theory of fun [79]. This approach was extended to cater for different player personas, each with their own fun formulation based on the divergence of their gameplay [82]; in a recent user study with 90 participants, this fun metric was found to be consistent with viewer expectations [83].

A notable family of theory-based affect comes from the *Temporal Difference Reinforcement Learning* (TDRL) theory of emotion [84], which simulates agents' emotional responses from the RL framework itself. Emotions are approximated using temporal difference (TD) errors: distress arises from negative TD errors, and joy from positive ones. Anticipatory emotions such as fear and hope are modeled using predictions from a forward model, representing the expected TD error. Regret corresponds to the negative difference between expected and actual rewards [85], meaning an agent has taken an action that gave it a poorer outcome than expected. Rather than serving as direct surrogates for human emotions, TDRL emotions aim to enhance transparency in agent behavior [86] and foster improved human-agent alignment and collaboration [87] by simulating its own internal emotional reactions. In contrast to the other ARL methods in this section, TDRL emotions do not act as the reward signal for training the agent, but are derived from behavioral rewards to explain the internal state of the agent. Therefore, the approximation of emotions based on TD errors can lead to more believable and transparent agent behaviors, as validated in a robot object detection task [88] and in a grid-based navigation task [89]. Moreover, concepts such as simulated regret can lead to improved agent performance: Soman et al. [90] used simulated regret to adaptively balance exploration and exploitation through an ϵ-greedy action selection strategy. High regret indicated a poor model of the reward landscape and caused exploration to increase. This is especially useful in dynamic environments, where agents can adapt their behavior based on fluctuating regret levels.

Zhang et al. [65] developed an agent that combines appraisal theory with TDRL emotions by incorporating four key appraisal dimensions (i.e. suddenness, goal relevance, conduciveness, and power) directly into the RL learning loop. These appraisals are classified into modal emotions (happiness, boredom, and irritation) using a Support Vector Machine (SVM) classifier calibrated on simulated data, with emotional states persisting across time steps. The agent is assigned a positive or negative reward depending on whether they correctly enter the desired goal state. Validation experiments with 72 human participants in reading comprehension tasks showed a strong alignment between model predictions and self-reported emotions. Zhang et al. [91] extended this work by integrating Scherer's Component Process Model [92] with RL to predict emotional responses in interactive task environments (see Fig. 4). Similar to their previous work, they perform appraisal computations from the agent's TD updates, and employ SVM classifiers trained on theory-defined appraisal patterns to predict intensities of modal emotions, including joy, fear, shame, and desperation. Validation with human participants, who were tasked with reading technology interaction scenarios, showed that the model successfully captured individual differences in emotional responses. Finally, Prasad et al. [93] also use appraisal theory to drive learning in a PPO agent by introducing cognitive appraisal variables—such as certainty, novelty, and goal congruence—into various reward functions, and were shown to successfully simulate mental health disorder behaviors such as anxiety and OCD.

IV. Affective Actions

In this section, we discuss three major categories of action spaces used by ARL agents. First, we cover actions that

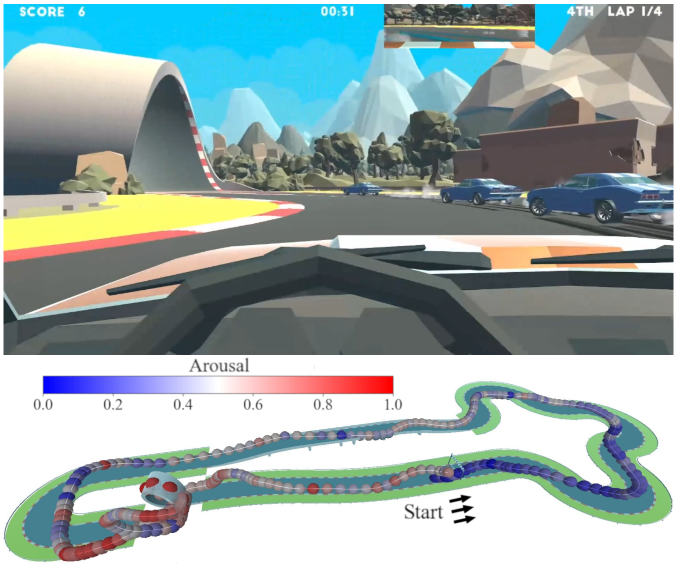


Fig. 5. An example play trace in the *Solid Rally* racing game (top image) taken from [52], where an RL agent was trained to imitate the average behavior and arousal of expert players in the AGAIN dataset [94]. The depicted trace in the bottom image is colored according to the arousal responses generated by the agent during gameplay;blue and red color indicates low and high arousal, respectively.

affect the *behavior* of the agent (see Section IV-A). We then move on to survey agents that learn to take actions that alter the *content* of their environment (see Section IV-B). Finally, we cover agents taking actions that directly alter their internal *affective state* rather than any outward behavior or environmental context (see Section IV-C).

A. Behavior

An RL agent typically learns by experience to take actions within a fixed set of contexts or interactions provided by an environment; we refer to this sequential decision-making as agent *behavior*. As covered in Section III, ARL diverges from typical RL research in that the agent learns to exhibit behaviors which elicit specific affective responses from a human or affect model, rather than a reward signal provided by the environment to accomplish a specific task. In this section, we illustrate the various forms of behavior actions that are currently used in ARL.

Games offer perhaps the most challenging form of sequential decision making, as an agent typically is required to perform fine-grained, fast-paced, and precise actions. In most real-time games, these actions are related to pathfinding or (short or long-term) planning, such as learning to navigate within a level and reach a target [20], eliminate enemies [53], or race against other drivers [51]. Some games also offer a rich social interaction setting, allowing RL agents to be trained to converse with other players [57]. Affect is typically used to take more human-like behavioral actions [49], [51], [53], or as a learning signal for optimizing action selection [20], [52]. Figure 5 depicts an indicative example of a gameplaying agent which learns to drive whilst imitating the behavior and affective responses of players.

Robotics presents substantial similarities with video games in terms of the agent's behavioral space. In robotics, an RL agent learns to control the physical locomotion of the robot [30]. For example, Kim et al. [42] trained a robotic arm to move physical items in a cluttered environment more safely using ErrP signals from a human-in-the-loop. Another interesting example by Churamani et al. [57] sits at the intersection of robotics and video games: they trained a physical robot to negotiate the value of an exchange with a human while playing the *Ultimatum* social game, and used a dynamic internal affective mood to guide its decision making during negotiation. Similarly, Angelopoulos et al. [95] used human feedback to train a robot to play the game *Mastermind*, where the robot must learn to guess the code provided by a human player by pointing towards colored balls. They then used TDRL emotions (see Section III-D) derived from the robot's actions and uncertainty to improve the transparency of its behavior to the human teacher.

Intuitively, actions which govern agent behavior are commonly used for RL agents in autonomous driving systems [96]. However, in practice, ARL has so far focused solely on recognition of the driver's mental state for safety applications such as maintaining driver alertness [37], rather than physical control of the vehicle. An indicative study of this approach by Yousaf et al. [38] trained a deep RL agent to take corrective actions based on EEG signals of the driver. Such actions included auditory alerts and speed adjustment to promote better driver attention.

In healthcare, this type of action typically require the agent to interact with a human patient, either by providing direct care or by alerting other staff for aid [97]. An indicative example of this use case in ARL can be seen in [72], where an LLM-powered agent generates more natural verbal assistance for patients living with dementia. Shaik et al. [48] trained RL agents to monitor simulated patients' physiological signals (i.e., heart rate, respiratory rate, etc.) and alert the appropriate medical emergency team for escalating care if required.

Finally, ARL can be used for general social interaction settings, such as open conversations with a human or another agent. The only indicative example found from our literature survey is by Jhan et al. [62] who use RL for text generation by training a deep RL agent to process BERT encodings and retrieve the most empathic response in the EmpathicDialogues dataset [61], outperforming a generative agent that relies on a GPT model. Other studies in the social interaction domain tend to use RL for finetuning transformer models [63], [98] using emotion feedback; those are detailed further in Section IV-C.

B. Context

The next type of action we survey is related to *context*, by which the RL agent takes actions that generate new contexts, or offer new contextual interactions in the environment. As opposed to behavioral agents, which operate within a fixed set of constraints provided by the environment, context agents attempt to create entirely new contexts and interactions according to a desired emotional experience for humans or

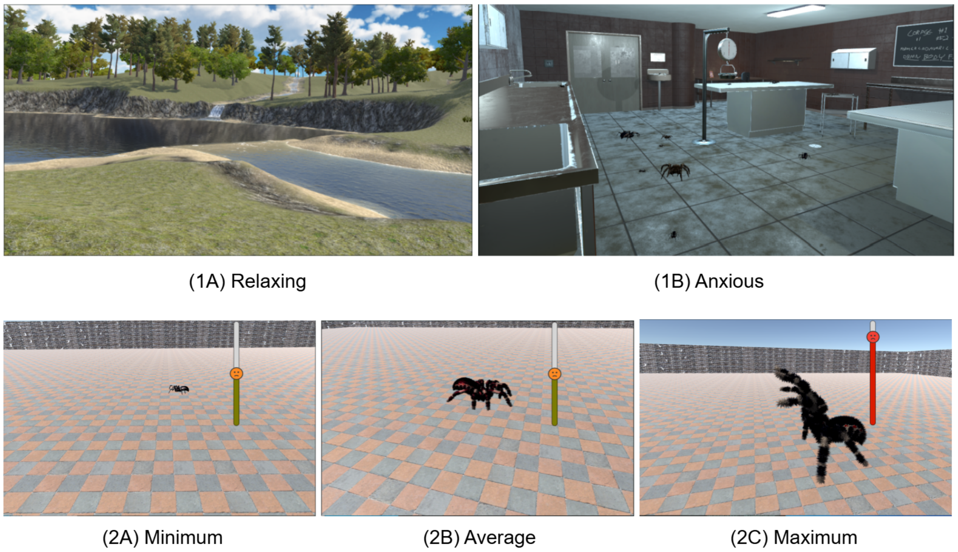

(1A) Relaxing
(1B) Anxious
(2A) Minimum
(2B) Average
(2C) Maximum

Fig. 6. Examples of personalized arachnophobia therapy contexts generated using EDRL. The top row depicts the two types of VR environments developed—a relaxing natural environment and a stressful setting—whilst the bottom row depicts the various levels of targeted anxiety for the generated spiders and their attributes. Visual taken from [19], [99] with authors' permission.

other agents. This makes context actions particularly useful in domains with a strong emphasis on personalized interactions between agents, humans, and their environment. In this section, we survey studies on context adaptation—as a sequence of actions generated by an ARL agent—across different domains.

Designing new contexts for games based on the player's experience is considered one of the ultimate goals of AI-driven game design [5]. An indicative study of context-based ARL in games [18] generates endless *Super Mario Bros* (Nintendo, 1985) levels through an RL designer agent. The *Experience-Driven RL* (EDRL) [82] framework trains a level designer to take actions—which involve modifying level parameters, such as platform placement, enemy distribution, and power-up locations—to maximize a player's fun; the fun reward function is based on Koster's theory of fun [79] and is cross-verified against human players [83]. In a further development of this approach, Barthet et al. [100] used the same EDRL approach to train a designer agent that generates racetrack layouts by iteratively placing predefined track components (straights, curves, and bridges, etc.) so that it elicits specific arousal responses across different players.

Healthcare is another promising domain for personalized interaction and stimulus design, as the needs of individual patients are often critical when implementing a treatment plan. An indicative proof of concept study illustrating this idea is by Mahmoudi-Nejad et al. [19], [99], [101], who implemented a closed-loop system for personalized exposure therapy as a treatment approach against arachnophobia. In their system, the ARL agent takes design actions that incrementally adjust attributes of virtual spiders to maintain optimal anxiety levels for therapeutic exposure. The designer agent uses Tabular Q-Learning to generate attributes, which correspond to the spider's locomotion, range of movement, closeness to the patient, hairiness, size, and color, examples of which can be seen in Fig. 6. Shen et al. [102] also propose a closed-loop system, where the agent's actions involve generating personalized therapeutic music conditioned on real-time EEG feedback, which the authors aim to evaluate in the near future. The action space encompasses musical parameters such as tempo, key, and instrumentation, with rewards computed from both music-theoretic quality metrics and observed emotional state changes in users. Sharma et al. [67] trained agents to suggest edits to sentences to increase empathy for support-providers in conversations with support-seekers whilst maintaining coherence, fluency, diversity, and adherence to the context.

C. Internal Affect

Beyond behavior and contexts, ARL agents may learn to make internal affect predictions according to observations received from the environment. Naturally, this approach can train emotion recognition models similarly to supervised learning: the agent is rewarded for making predictions that align with a ground truth. Studies using this approach generally do not focus on domain-specific environments. rather they test ARL's effectiveness across generic affect corpora in an offline manner. We categorize this type of action into two groups: (a) models trained from scratch using traditional RL methods such as DQN and PPO, and (b) models fine-tuned from existing pre-trained models, such as LLMs.

Zhou et al.'s [103] Emotion-Agent exemplifies agents using traditional RL algorithms. They train a PPO agent to identify emotionally salient neural activity on provided EEG signals. The agent learns to classify segments based on their emotional relevance, with rewards based on proximity to emotion cluster centers. Another notable example is the speech emotion recognition work by Rajapakshe et al. [104], where the actions of a deep RL agent correspond to predicting emotional states from speech features. The agent learns a policy that maps acoustic features to emotion categories, receiving rewards based on classification accuracy. In their follow-up study, Rajapakshe et al. [105] performed domain adaptation using deep Q-learning over an existing pretrained speech emotion recognition model. Tangentially, in the early anger detection system by Lakomkin et al. [54], the REINFORCE algorithm was used to train agents that predict the onset of anger in speech, with actions representing confidence thresholds for making emotional predictions with an existing SL affect model. Finally, Rahman et al. [71] train a dueling DQN agent for predicted emotions ranging from happiness to frustration, with a fixed positive or negative reward for correct or incorrect classifications.

LLMs' ability to handle new modalities has been applied to predict affect labels in existing corpora [106], [107]. Similar to RLHF [16], a recent study [98] explored the use of Reinforcement Learning with Verifiable Rewards (RLVF)—first introduced by DeepSeek R1 [108]—to fine-tune LLMs for improved emotion prediction. Specifically, they finetuned the HumanOmni-0.5B model on 232 samples from the EMER dataset and 348 samples from their own manually annotated HumanOmni dataset. Their reward is broken down into two parts: an accuracy reward, which rewards the model for identifying the correct emotion compared to the ground truth label, and a format reward to ensure the model sticks to a strict output format that is interpretable for the downstream analysis. Results show that an LLM trained with this extension of RLVF outperforms three baseline models significantly across three different video affect corpora, highlighting the strength of RL for improving emotion detection in LLMs.

V. Affective States

The aim of the ARL agent within our framework is to learn a mapping between its current state and a desired affective action. Therefore, the state space must embed the necessary affective and contextual information required for the agent to learn such a mapping. In this section, we survey the different state spaces used in ARL methods and categorize them into three main types. First, we cover agents which rely on features extracted from *verbal* observations of humans such as text and speech (Section V-A). Next, we survey *non-verbal* cues from humans such as facial expression, physiological signals, and body stance (Section V-B). Finally, Section V-C covers ARL agents that use features extracted from contextual information available in the environment.

A. Verbal Observations

Predicting affect through human verbal communication, such as text and speech, is a popular line of research within affective computing [109]. Verbal information can potentially offer rich sources of both semantic content and implicit emotional states of the speaker. These observations are becoming increasingly popular to use in affect modeling, especially with the emergence of the transformer architecture and the resulting improvement in natural language processing capabilities. Affective computing work that relies on verbal cues mostly tackles emotion prediction and conversation. Therefore, we organize this section across the two types of verbal observations available: text and speech.

For text-based ARL, an indicative early example by Li et al. [56] feeds text from emotional conversations from the NLPCC2017 dataset into a constrained encoder-decoder architecture to generate more relevant replies. In recent studies, the transformer architecture has become the dominant underlying method employed for automated feature extraction. For example, Shin et al. [59] take raw conversation text and use a combination of a fine-tuned BERT model for sentiment classification, and embeddings extracted using Word2Vec [110] to generate more empathic responses. Similarly, Sharma et al. [67] use raw text posts from support seekers and responders on a mental health platform to generate sentence edits to improve empathy using a GPT-2 architecture. For example, the NARLE framework [63] uses a method called *ScopeIt* [111] for filtering task-relevant sentences from text that express emotion through an architecture using BERT for more concise observations.

For speech-based ARL, a common approach for the state representation relies on extracted acoustic features such as MFCCs [112], given their proven effectiveness for the task in the broader field of AC. For example, Rajapakshe et al. [104] used a CNN-LSTM architecture to process MFCC features, which act as the state space for an RL agent which achieved improved accuracy on standard emotion recognition benchmarks such as IEMOCAP and SAVEE. Lakomkin et al. [54] also extracted a fixed set of 15 MFCC-based features for a real-time emotion classification agent in conversations from the IEMOCAP corpus. In contrast, Liu et al. [58] forgo the use of MFCCs and instead use an audio encoder for extracting emotion embeddings from a raw reference speech segment, which is combined with a text encoder for a more emotionally appropriate text-to-speech model.

B. Non-Verbal Observations

Non-Verbal cues from humans-in-the-loop—such as facial or bodily reactions—are one of the most common forms of observation for visual affect sensing in computer vision [113]. These are particularly useful for tasks which place more emphasis on physical or physiological cues from humans-in-the-loop—rather than verbal interactions between two parties. Such cues are useful for healthcare [48] and robotics [76], which we cover below.

In the domain of affective robots, the ARL agent is typically deployed alongside a human-in-the-loop, and therefore combines sensors governing its own state with the predicted state of its human counterparts. For example, Zhu et al. [76] developed a bipedal robot that uses projected pedestrian position and velocities alongside other non-verbal cues to navigate roads more safely. Churumani et al. [57] employed pretrained facial emotion recognition models to extract features from video streams. These typically produce either categorical emotion labels or continuous valence-arousal values that serve as part of the agent's observation space.

In healthcare applications, there is an emphasis on capturing non-verbal features that describe the emotional state of any human patient involved. Physiological observations are commonly used in patient care ARL agents, such as the system by Shaik et al. [48], which uses the patient's heart and

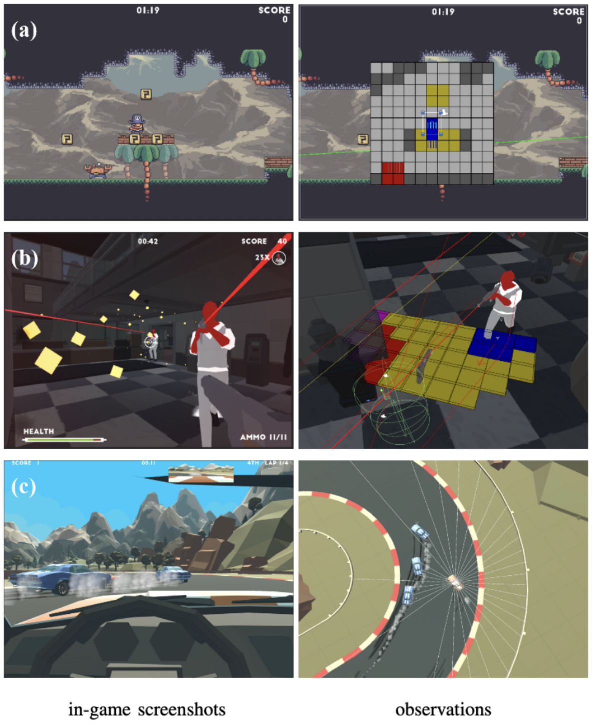


Fig. 7. Examples of contextual state spaces, visualized alongside the games from the *Affectively Framework* [53]. From top to bottom: (a) *Pirates* platformer game, (b) *Heist* first-person shooter, and (c) *Solid Rally* racing game.

respiration rate and their skin temperature as its state space. Duttu et al. [114] include both musical features (i.e., tempo, key, timbre) and the user's EEG responses to music to generate music for emotional management therapy. In dementia care scenarios [72], camera-based systems could track patient facial expressions and body language to assess confusion, distress, or engagement levels. These visual cues complement other observation modalities to provide a comprehensive view of the patient's state.

C. Contextual Observations

As opposed to the human observation spaces described above, contextual observations encode information specific to the environment, the agent, and the task at hand. Typically, these features rely on domain knowledge provided by the human designer when implementing the environment and the agents to be trained. In ARL, these features are chosen with an emphasis on capturing both functionally and emotionally relevant information for the agent's affective learning compared to traditional RL, where functional information is the most important. In this section, we group the different forms of contextual observations according to the domain tackled by the environment.

Within the domain of games, the type of contextual observations used depends heavily on the genre of the game being played, and often overlaps significantly with typical RL methods. For example, Barthet el al. [53] trained affective gameplaying agents across three different game genres, each with its own contextual observations. For racing games, such observations include the agent's velocity, rotation, whether it is off-road or crashing, and the distances to its nearest surroundings [51], [52]. For platformer games such as *Super Mario Bros* (Nintendo, 1985), these observations may include the agent's current velocity and health, whether it's currently in the air due to a jump, or how many collectibles it has picked up [18], [53]. For first-person shooters, relevant contextual information might include the amount of ammunition and enemies left, the number of enemies in line of sight, and the distance and angle to the nearest enemy [53].

Robotics relies on contextual observations that govern an agent's internal state through specialized sensors, which capture its position, movement, and relation to other objects. An illustrative study by Kim et al. [44] trained a robotic arm to physically wave in response to prosodic feedback from a human participant, using a state space representing the locomotive state–i.e., the current waving motion—for learning. In another example study [95], an ARL robot was trained to play the board game *Mastermind* using features that describe the secret codes guessed by the agent so far; the robot physically pointed to colored balls on a board.

Within other domains such as healthcare and autonomous driving, there is an expected focus on human (verbal or not) over contextual observations, given the nature of these domains. In a tangential study by Deepa et al. [115], the ARL state considers the car's behavior alongside physiological cues obtained from the driver to predict the driver's emotional state; however, the specifics of the car's observations are not described in detail.

VI. Open Problems and Outlook

This survey has clearly illustrated the diversity of approaches within ARL, both in terms of how affect is used within the RL process and affective loop, and in terms of the tasks and domains tackled during learning. As this area matures, we expect research on ARL to contribute both to improving the current challenges of typical RL agents, as well as help advance the field of affective computing by providing an alternative to the supervised learning paradigm. In this section, we outline the current issues in ARL training, and discuss our outlook with respect to the domains tackled—or yet to be explored—by this research area, as illustrated in Fig. 8.

Similar to other research areas within affective computing, the main challenge in ARL is producing an accurate and reliable affective signal for training. A large portion of the works surveyed in this paper rely on either a human-in-the-loop or model-driven methods to derive an appropriate affective reward signal. This typically requires sensing through physiological signals or less obtrusive methods, such as manual annotation and facial and body expression recognition through cameras. Depending on the domain, a reliance on sensors can limit feasibility when deploying outside of the lab, where data is less likely to be clean and within distribution. Reliably approximating affective states with less reliance on intrusive sensors will be a significant step forward for this line of research to deploy more easily outside of controlled lab environments.

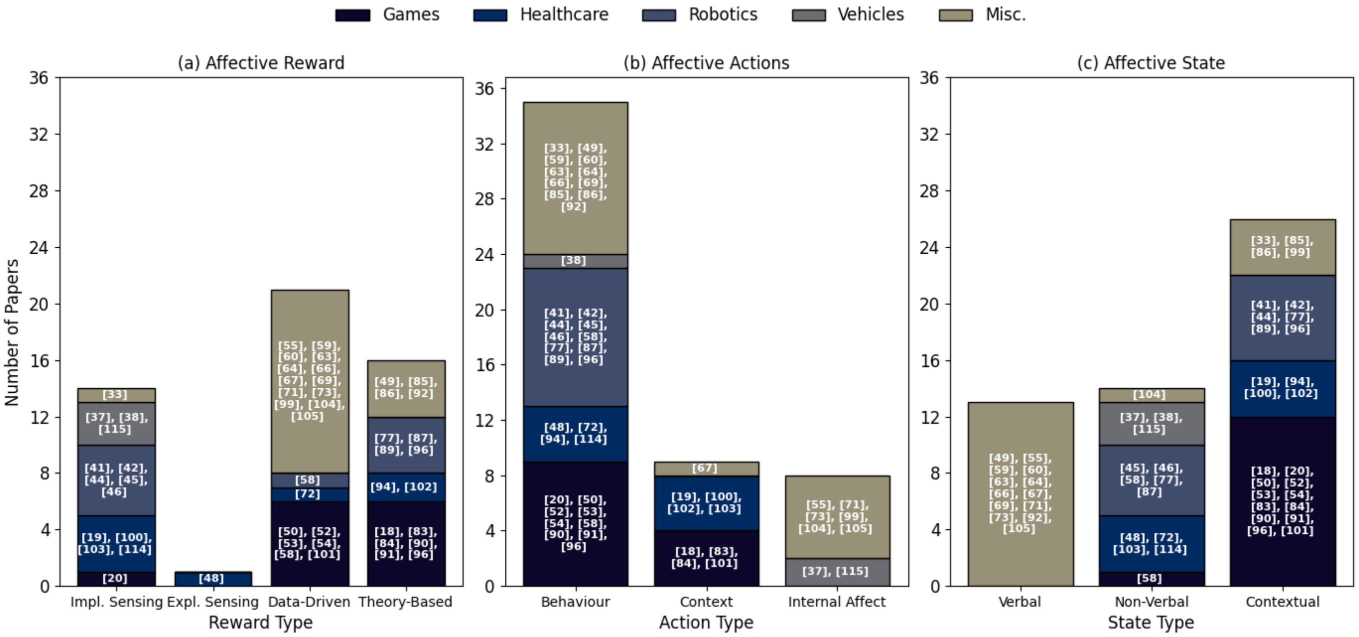


Fig. 8. Distribution of papers identified in our survey across (a) affective rewards, (b) affective actions, and (c) affective states, further grouped by the five major domains used in ARL. The miscellaneous domain refers to ARL papers that do not apply to a particular domain but use RL for tasks such as emotion recognition [54], [104] or training affective conversation agents [63], [67].

Digital games have a strong presence within ARL and present a diverse range of complexities in terms of the task and form of agent interaction. This is because games fit naturally with RL research, provide a rich form of human-computer interaction [5], and are a safe and fast environment for training. ARL within games can be grouped according to the ultimate goal of the agent. The first grouping uses affect as an implicit reward signal for learning optimal behavior through physiological signals—such as ErrP signals [39]—without using task-specific reward functions. The survey highlights how such signals can generalize well across similar, simpler games; however, the state of the art is yet to tackle more complex and vast games that are more typical of modern commercial games. The second grouping uses model-based methods through theory or data-driven methods. As these do not require a human-in-the-loop, the barrier for entry for this line of research can be significantly lower and more feasible for applications outside of the lab. However, more work needs to be done to release publicly available affective corpora paired with appropriate RL environments (such as [53]), which can fuel future research. With the advancements in computer vision capabilities, we believe vision-based affective game corpora (such as [116], [117]) present a strong opportunity for testing ARL on more complex, AAA games where access to the game engine is not available.

Healthcare arguably presents the most impactful domain for ARL research, as it can help ease the burden on overburdened healthcare networks, which remain a challenge worldwide. Within ARL, there is a clear focus on adaptable agents that can provide care to patients suffering from physical or mental issues, such as phobias [19], dementia [72], and emergency care [48]. Given the human-centric nature of healthcare, it is no surprise that the most common method used in training follows a human-in-the-loop architecture with physiological signals to derive reward functions. With the advancements of AI in healthcare, such as SL for interpreting scans [118], future research should work alongside healthcare staff in patient care interactions, where ARL's strengths show compared to other AI methods. By including the affective states of the patient and healthcare workers, ARL healthcare agents can take more informed decisions that improve both physical and mental well-being.

Robotics was a very popular domain for early ARL research, using simulated agent emotions [14]. In our survey, robotics has been a popular testbed for implicit feedback rewards [35], [41] or theory-based affect reward functions [86], [95]. Affective robotics presents a strong domain for research where there is more emphasis on physical human-agent collaboration, especially in overlap with other domains such as healthcare, where collaboration between doctors and surgeons with machines is becoming more prevalent in modern medicine [119]. Such agents could provide faster responses (e.g., by alerting staff) [120] and manage emergency situations (e.g., deliver CPR) [121] until caregivers are available. This capability can help reduce stress for both carers and patients during vulnerable periods—such as overnight hours when fewer medical staff are present and fatigue may more strongly affect decision-making.

In vehicular applications, there is a focus on detecting human drivers' affective state and reacting accordingly to maximize alertness and safety on the road. In the current state of the art, the agents typically react through the car's audio system—such as playing a beep or music—or by altering the car's climate control system. Modern vehicles can provide feedback to the steering wheel for a haptic feedback option. With the rapid development of in-car sensors and systems (e.g.

devices that can read biometric signals from the driver with better accuracy), we expect this line of research to see further development and deployment into real-world scenarios.

As shown in Fig. 8, a substantial body of work within ARL research does not apply itself to a specific domain but either builds agents for emotion recognition tasks [54], [104], [105], or trains conversation agents with an emphasis on affective replies such as empathy [67]. Going forward, this type of research can greatly benefit other, more domain-focused work in ARL by creating better affect models more suited for interactive settings and continued, incremental learning compared to SL methods. This would allow for a more dynamic system which better closes the affective loop, as depicted in Section II-B. This would, for example, allow game designers to create games which continually adapt to their players' affective preferences as they evolve over time. Healthcare agents could also benefit by continually improving their models of patients' affective state throughout their care, as opposed to using a fixed model.

There are many other domains involving interaction between humans and computers or other devices that are yet to be explored in ARL research. For example, ARL can be used to adjust the interfaces of websites and apps, such as social media, based on the real-time emotional experience of the user. With the rise of smart devices in homes, a similar idea to the study applied in vehicles [38] can be applied to maximize comfort and positive affect through climate control, music playback, and lighting in homes. We believe ARL is a natural fit for any such affective interactions where there is a strong emphasis on sequential and real-time decision making without the need for collecting large amounts of human data.

As with any research on affective computing there must be a strong emphasis on ethics and ensuring agents do not cause harm when deployed. Currently, there is an overwhelming focus in ARL literature on promoting positive affect, however such systems could easily be trained by malicious actors to do the opposite and cause harm. Furthermore, even positive intentions can cause unwanted harm to users [122], such as creating echo chambers or be manipulated to promote dangerous ideas. As a result, ARL research must also focus on transparency, safety, and controllability to ensure that it is ultimately beneficial to society.

VII. Conclusion

Affective reinforcement learning has clear promise to push the development of affect models outside of the typical supervised learning paradigm followed within affective computing. The area, whilst still relatively small compared to other paradigms in affective computing, is seeing more and more promising research and is inching closer to the ultimate goal of closed-loop affective computing systems. For instance, affective adaptation of stimuli based on recognized behaviors and emotions is becoming more common, with successful applications both in games and in healthcare. As RL algorithms and traditional affect models continue to develop, one can expect ARL to gain momentum and see widespread real-world deployment and adoption.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[2] R. W. Picard, *Affective computing*. MIT press, 2000.

[3] C. Yu, J. Liu, S. Nemati, and G. Yin, "Reinforcement learning in healthcare: A survey," *ACM Computing Surveys (CSUR)*, vol. 55, no. 1, pp. 1–36, 2021.

[4] S. Zepf, J. Hernandez, A. Schmitt, W. Minker, and R. W. Picard, "Driver emotion recognition for intelligent vehicles: A survey," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–30, 2020.

[5] G. N. Yannakakis and D. Melhart, "Affective game computing: A survey," *Proceedings of the IEEE*, vol. 111, no. 10, pp. 1423–1444, 2023.

[6] K. Höök, "Affective loop experiences–what are they?," in *Proceedings of the third international conference on Persuasive Technology*, Springer, 2008.

[7] G. Pei, H. Li, Y. Lu, Y. Wang, S. Hua, and T. Li, "Affective computing: Recent advances, challenges, and future trends," *Intelligent Computing*, vol. 3, p. 0076, 2024.

[8] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang, *et al.*, "A systematic review on affective computing: Emotion models, databases, and recent advances," *Information Fusion*, vol. 83, pp. 19–52, 2022.

[9] R. S. Sutton, A. G. Barto, *et al.*, *Reinforcement learning: An introduction*, vol. 7. MIT press Cambridge, 1992.

[10] A. Ortony, G. L. Clore, and A. Collins, *The cognitive structure of emotions*. Cambridge university press, 2022.

[11] N. Akalin and A. Loutfi, "Reinforcement learning approaches in social robotics," *Sensors*, vol. 21, no. 4, p. 1292, 2021.

[12] J. Hoey, T. Schröder, and A. Alhothali, "Affect control processes: Intelligent affective interaction using a partially observable markov decision process," *Artificial Intelligence*, vol. 230, pp. 134–172, 2016.

[13] F. Kaptein, J. Broekens, K. Hindriks, and M. Neerincx, "The role of emotion in self-explanations by cognitive agents," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 88–93, IEEE, 2017.

[14] T. M. Moerland, J. Broekens, and C. M. Jonker, "Emotion in reinforcement learning agents and robots: a survey," *Machine Learning*, vol. 107, pp. 443–480, 2018.

[15] A. Paiva, I. Leite, H. Boukricha, and I. Wachsmuth, "Empathy in virtual agents and robots: A survey," *Transactions on Interactive Intelligent Systems (TiiS)*, vol. 7, no. 3, pp. 1–40, 2017.

[16] T. Kaufmann, P. Weng, V. Bengs, and E. Hüllermeier, "A survey of reinforcement learning from human feedback," *Transactions on Machine Learning Research*, 2025. Survey Certification.

[17] P. Jemioło, D. Storman, M. Mamica, M. Szymkowski, W. Żabicka, M. Wojtaszek-Główka, and A. Ligkeza, "Datasets for automated affect and emotion recognition from cardiovascular signals using artificial intelligence—a systematic review," *Sensors*, vol. 22, no. 7, p. 2538, 2022.

[18] T. Shu, J. Liu, and G. N. Yannakakis, "Experience-driven pcg via reinforcement learning: A super mario bros study," in *Proceedings of the International Conference on Games (CoG)*, IEEE, 2021.

[19] A. Mahmoudi-Nejad, M. Guzdial, and P. Boulanger, "Personalizing exposure therapy via reinforcement learning," *arXiv preprint arXiv:2504.14095*, 2025.

[20] D. Xu, M. Agarwal, E. Gupta, F. Fekri, and R. Sivakumar, "Accelerating reinforcement learning using eeg-based implicit human feedback," *Neurocomputing*, vol. 460, pp. 139–153, 2021.

[21] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, pp. 279–292, 1992.

[22] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine learning*, vol. 8, pp. 229–256, 1992.

[23] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proceedings of the International conference on machine learning*, pp. 1928–1937, PmLR, 2016.

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

[25] C. Yu, A. Velu, E. Vinitsky, J. Gao, Y. Wang, A. Bayen, and Y. Wu, "The surprising effectiveness of ppo in cooperative multi-agent games," *Advances in neural information processing systems*, vol. 35, pp. 24611–24624, 2022.

[26] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-baselines3: Reliable reinforcement learning implementations," *Journal of machine learning research*, vol. 22, no. 268, pp. 1–8, 2021.

[27] A. Ecoffet, J. Huizinga, J. Lehman, K. O. Stanley, and J. Clune, "First return, then explore," *Nature*, vol. 590, no. 7847, pp. 580–586, 2021.

[28] D. Pathak, P. Agrawal, A. A. Efros, and T. Darrell, "Curiosity-driven exploration by self-supervised prediction," in *Proceedings of the International conference on machine learning*, pp. 2778–2787, PMLR, 2017.

[29] B. Eysenbach, A. Gupta, J. Ibarz, and S. Levine, "Diversity is all you need: Learning skills without a reward function," *arXiv preprint arXiv:1802.06070*, 2018.

[30] B. Singh, R. Kumar, and V. P. Singh, "Reinforcement learning in robotic applications: a comprehensive survey," *Artificial Intelligence Review*, vol. 55, no. 2, pp. 945–990, 2022.

[31] E. Mosqueira-Rey, D. A.-R. Elena Hernández-Pereira1, J. Bobes-Bascarán1, and Ángel Fernández-Leal1, "Human-in-the-loop machine learning: a state of the art," *Artificial Intelligence Review*, vol. 56, p. 3005–3054, 2023.

[32] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza, "Power to the people: The role of humans in interactive machine learning," *AI Magazine*, vol. 35, no. 4, pp. 105–120, 2014.

[33] R. Chavarriaga and J. d. R. Millán, "Learning from eeg error-related potentials in noninvasive brain-computer interfaces," *Transactions on neural systems and rehabilitation engineering*, vol. 18, no. 4, pp. 381–388, 2010.

[34] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[35] I. Iturrate, L. Montesano, and J. Minguez, "Robot reinforcement learning using eeg-based reward signals," in *Proceedings of the IEEE international conference on robotics and automation*, pp. 4822–4829, IEEE, 2010.

[36] J. J. Shih, D. J. Krusienski, and J. R. Wolpaw, "Brain-computer interfaces in medicine," *Mayo clinic proceedings*, vol. 87, no. 3, pp. 268–279, 2012.

[37] Y. Ming, D. Wu, Y.-K. Wang, Y. Shi, and C.-T. Lin, "Eeg-based drowsiness estimation for driving safety using deep q-learning," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 4, pp. 583–594, 2020.

[38] M. Yousaf, M. Farhan, Y. Saeed, M. J. Iqbal, F. Ullah, and G. Srivastava, "Enhancing driver attention and road safety through eeg-informed deep reinforcement learning and soft computing," *Applied Soft Computing*, vol. 167, p. 112320, 2024.

[39] P. W. Ferrez and J. d. R. Millán, "Error-related eeg potentials generated during simulated brain–computer interaction," *Transactions on biomedical engineering*, vol. 55, no. 3, pp. 923–929, 2008.

[40] A. F. Salazar-Gomez, J. DelPreto, S. Gil, F. H. Guenther, and D. Rus, "Correcting robot mistakes in real time using eeg signals," in *2017 IEEE international conference on robotics and automation (ICRA)*, pp. 6570–6577, IEEE, 2017.

[41] I. Akinola, Z. Wang, J. Shi, X. He, P. Lapborisuth, J. Xu, D. Watkins-Valls, P. Sajda, and P. Allen, "Accelerated robot learning via human brain signals," in *Proceedings of the international conference on robotics and automation*, pp. 3799–3805, IEEE, 2020.

[42] S. Kim, H.-B. Shin, and S.-W. Lee, "Aligning humans and robots via reinforcement learning from implicit human feedback," *arXiv preprint arXiv:2507.13171*, 2025.

[43] G. Wu, G. Liu, and M. Hao, "The analysis of emotion recognition from gsr based on pso," in *Proceedings of the International symposium on intelligence information processing and trusted computing*, pp. 360–363, IEEE, 2010.

[44] E. S. Kim and B. Scassellati, "Learning to refine behavior using prosodic feedback," in *Proceedings of the 6th International Conference on Development and Learning*, pp. 205–210, IEEE, 2007.

[45] K. Weber, H. Ritschel, I. Aslan, F. Lingenfelser, and E. André, "How to shape the humor of a robot-social behavior adaptation based on reinforcement learning," in *Proceedings of the 20th international conference on multimodal interaction*, pp. 154–162, ACM, 2018.

[46] M. Pollak, A. Salfinger, and K. A. Hummel, "Teaching drones on the fly: Can emotional feedback serve as learning signal for training artificial agents?," *arXiv preprint arXiv:2202.09634*, 2022.

[47] J. Pérez, E. Dapena, and J. Aguilar, "Emotions as implicit feedback for adapting difficulty in tutoring systems based on reinforcement learning," *Education and Information Technologies*, vol. 29, no. 16, pp. 21015–21043, 2024.

[48] T. Shaik, X. Tao, L. Li, H. Xie, H.-N. Dai, F. Zhao, and J. Yong, "Adaptive multi-agent deep reinforcement learning for timely healthcare interventions," *arXiv preprint arXiv:2309.10980*, 2023.

[49] M. Barthet, A. Liapis, and G. N. Yannakakis, "Go-blend behavior and affect," in *Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, IEEE, 2021.

[50] D. Melhart, A. Liapis, and G. N. Yannakakis, "The Arousal video Game AnnotatIoN (AGAIN) dataset," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, 2022.

[51] M. Barthet, A. Khalifa, A. Liapis, and G. Yannakakis, "Generative personas that behave and experience like humans," in *Proceedings of the 17th International Conference on the Foundations of Digital Games*, ACM, 2022.

[52] M. Barthet, A. Khalifa, A. Liapis, and G. N. Yannakakis, "Play with emotion: Affect-driven reinforcement learning," in *Proceedings of the 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*, IEEE, 2022.

[53] M. Barthet, R. Gallotta, A. Khalifa, A. Liapis, and G. N. Yannakakis, "Affectively framework: Towards human-like affect-based agents," *Proceedings of the 12th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 2024.

[54] E. Lakomkin, M. A. Zamani, C. Weber, S. Magg, and S. Wermter, "Emorl: continuous acoustic emotion classification using deep reinforcement learning," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pp. 4445–4450, IEEE, 2018.

[55] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "Iemocap: Interactive emotional dyadic motion capture database," *Language resources and evaluation*, vol. 42, pp. 335–359, 2008.

[56] J. Li, X. Sun, X. Wei, C. Li, and J. Tao, "Reinforcement learning based emotional editing constraint conversation generation," *arXiv preprint arXiv:1904.08061*, 2019.

[57] N. Churamani, P. Barros, H. Gunes, and S. Wermter, "Affect-driven modelling of robot personality for collaborative human-robot interactions," *arXiv preprint arXiv:2010.07221*, 2020.

[58] R. Liu, B. Sisman, and H. Li, "Reinforcement learning for emotional text-to-speech synthesis with improved emotion discriminability," in *Proceedings of the Interspeech Conference*, pp. 4648–4652, 2021.

[59] J. Shin, P. Xu, A. Madotto, and P. Fung, "Generating empathetic responses by looking ahead the user's sentiment," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7989–7993, IEEE, 2020.

[60] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the conference on empirical methods in natural language processing*, pp. 1631–1642, 2013.

[61] H. Rashkin, E. M. Smith, M. Li, and Y.-L. Boureau, "Towards empathetic open-domain conversation models: A new benchmark and dataset," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2019.

[62] J.-H. Jhan, C.-P. Liu, S.-K. Jeng, and H.-Y. Lee, "Cheerbots: Chatbots toward empathy and emotionusing reinforcement learning," *arXiv preprint arXiv:2110.03949*, 2021.

[63] R. Zhou, S. Deshmukh, J. Greer, and C. Lee, "NaRLE: Natural language models using reinforcement learning with emotion feedback," *arXiv preprint arXiv:2110.02148*, 2021.

[64] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[65] J. E. Zhang, B. Hilpert, J. Broekens, and J. P. Jokinen, "Simulating emotions with an integrated computational model of appraisal and reinforcement learning," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024.

[66] F. Brahman and S. Chaturvedi, "Modeling protagonist emotions for emotion-aware storytelling," *arXiv preprint arXiv:2010.06822*, 2020.

[67] A. Sharma, I. W. Lin, A. S. Miner, D. C. Atkins, and T. Althoff, "Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach," in *Proceedings of the web conference*, pp. 194–205, 2021.

[68] A. Sharma, A. Miner, D. Atkins, and T. Althoff, "A computational approach to understanding empathy expressed in text-based mental health support," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2020.

[69] H. Ma, B. Zhang, B. Xu, J. Wang, H. Lin, and X. Sun, “Empathy level alignment via reinforcement learning for empathetic response generation,” *IEEE Transactions on Affective Computing*, 2025.

[70] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, “Exploring the limits of transfer learning with a unified text-to-text transformer,” *Journal of machine learning research*, vol. 21, no. 140, pp. 1–67, 2020.

[71] F. A. Rahman and G. Lu, “A contextualized real-time multimodal emotion recognition for conversational agents using graph convolutional networks in reinforcement learning,” *arXiv preprint arXiv:2310.18363*, 2023.

[72] F. Yuan, N. Hasnaeen, R. Zhang, B. Bible, J. R. Taylor, H. Qi, F. Yao, and X. Zhao, “Integrating reinforcement learning and ai agents for adaptive robotic interaction and assistance in dementia care,” *arXiv preprint arXiv:2501.17206*, 2025.

[73] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, “A survey of methods for explaining black box models,” *ACM computing surveys (CSUR)*, vol. 51, no. 5, pp. 1–42, 2018.

[74] P. Fung, D. Bertero, P. Xu, J. H. Park, C.-S. Wu, and A. Madotto, “Empathetic dialog systems,” in *Proceedings of the international conference on language resources and evaluation. European Language Resources Association*, 2018.

[75] P. Lopes, A. Liapis, and G. N. Yannakakis, “Targeting horror via level and soundscape generation,” in *Proceedings of the AAAI Artificial Intelligence for Interactive Digital Entertainment Conference*, 2015.

[76] W. Zhu, A. Raju, A. Shamsah, A. Wu, S. Hutchinson, and Y. Zhao, “Emobipednav: Emotion-aware social navigation for bipedal robots with deep reinforcement learning,” *arXiv preprint arXiv:2503.12538*, 2025.

[77] G. Ruggiero, F. Frassinetti, Y. Coello, M. Rapuano, A. S. di Cola, and T. Iachini, “The effect of facial expressions on peripersonal and interpersonal spaces,” *Psychological research*, vol. 81, no. 6, p. 1232–1240, 2017.

[78] G. N. Yannakakis and J. Togelius, *Artificial intelligence and games*, vol. 2. Springer, 2018.

[79] R. Koster, *Theory of fun for game design.* ” O’Reilly Media, Inc.”, 2013.

[80] M. Csikszentmihalyi, *Beyond boredom and anxiety.* Jossey-bass, 2000.

[81] G. N. Yannakakis and J. Hallam, “Capturing player enjoyment in computer games,” in *Advanced Intelligent Paradigms in Computer Games*, pp. 175–201, Springer, 2007.

[82] Z. Wang, J. Liu, and G. N. Yannakakis, “The fun facets of mario: Multifaceted experience-driven pcg via reinforcement learning,” in *Proceedings of the 17th International Conference on the Foundations of Digital Games*, pp. 1–8, 2022.

[83] Z. Wang, Y. Li, H. Du, J. Liu, and G. N. Yannakakis, “Fun as moderate divergence: Evaluating experience-driven pcg via rl,” *IEEE Transactions on Games*, 2024.

[84] J. Broekens, “A temporal difference reinforcement learning theory of emotion: unifying emotion, cognition and adaptive behavior,” *arXiv preprint arXiv:1807.08941*, 2018.

[85] J. Broekens and L. Dai, “A tdrl model for the emotion of regret,” in *Proceddings of the 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 150–156, IEEE, 2019.

[86] J. Broekens and M. Chetouani, “Towards transparent robot learning through tdrl-based emotional expressions,” *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 352–362, 2019.

[87] B. Hilpert, “Closing the teacher-learner loop: The role of affective signals in interactive rl,” in *Proceedings of the 12th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 97–101, IEEE, 2024.

[88] M. Matarese, S. Rossi, A. Sciutti, and F. Rea, “Towards transparency of TD-RL robotic systems with a human teacher,” *arXiv preprint arXiv:2005.05926*, 2020.

[89] L. Dai and J. Broekens, “Simulating fear as anticipation of temporal differences: an experimental investigation,” in *Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, IEEE, 2021.

[90] G. Soman, M. Judy, and S. Madria, “Regret emotion based reinforcement learning for path planning in autonomous agents,” in *Proceedings of the 12th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 266–274, IEEE, 2024.

[91] J. E. Zhang, J. Broekens, and J. Jokinen, “Modeling cognitive-affective processes with appraisal and reinforcement learning,” *IEEE Transactions on Affective Computing*, 2024.

[92] K. R. Scherer, “The dynamic architecture of emotion: Evidence for the component process model,” *Cognition and emotion*, vol. 23, no. 7, pp. 1307–1351, 2009.

[93] H. Prasad, C. Jacob, *et al.*, “Appraisal-guided proximal policy optimization: Modeling psychological disorders in dynamic grid world,” *arXiv preprint arXiv:2407.20383*, 2024.

[94] D. Melhart, A. Liapis, and G. N. Yannakakis, “The arousal video game annotation (again) dataset,” *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 2171–2184, 2022.

[95] G. Angelopoulos, A. Rossi, G. L’Arco, and S. Rossi, “Transparent interactive reinforcement learning using emotional behaviours,” in *Proceedings of the International Conference on Social Robotics*, pp. 300–311, Springer, 2022.

[96] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. Al Sallab, S. Yogamani, and P. Pérez, “Deep reinforcement learning for autonomous driving: A survey,” *IEEE transactions on intelligent transportation systems*, vol. 23, no. 6, pp. 4909–4926, 2021.

[97] G. N. Yannakakis, “Enhancing health care via affective computing,” *Malta Journal of Health Sciences*, vol. 5, no. 1, pp. 38–42, 2018.

[98] J. Zhao, X. Wei, and L. Bo, “R1-omni: Explainable omni-multimodal emotion recognition with reinforcement learning,” *arXiv preprint arXiv:2503.05379*, 2025.

[99] A. Mahmoudi-Nejad, M. Guzdial, and P. Boulanger, “Spiders based on anxiety: How reinforcement learning can deliver desired user experience in virtual reality personalized arachnophobia treatment,” *arXiv preprint arXiv:2409.17406*, 2024.

[100] M. Barthet, D. Branco, R. Gallotta, A. Khalifa, and G. N. Yannakakis, “Closing the affective loop via experience-driven reinforcement learning designers,” *Proceedings of the 12th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2024.

[101] A. Mahmoudi-Nejad, M. Guzdial, and P. Boulanger, “Arachnophobia exposure therapy using experience-driven procedural content generation via reinforcement learning (edpcgrl),” in *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 17, pp. 164–171, 2021.

[102] L. Shen, H. Zhang, C. Zhu, R. Li, K. Qian, W. Meng, F. Tian, B. Hu, B. W. Schuller, and Y. Yamamoto, “A first look at generative artificial intelligence based music therapy for mental disorders,” *IEEE Transactions on Consumer Electronics*, 2024.

[103] Z. Zhou, Q. Liu, J. Wang, and Z. Liang, “Emotion-agent: Unsupervised deep reinforcement learning with distribution-prototype reward for continuous emotional eeg analysis,” *arXiv preprint arXiv:2408.12121*, 2024.

[104] T. Rajapakshe, R. Rana, S. Khalifa, J. Liu, and B. Schuller, “A novel policy for pre-trained deep reinforcement learning for speech emotion recognition,” in *Proceedings of the Australasian Computer Science Week*, pp. 96–105, ACM, 2022.

[105] T. Rajapakshe, R. Rana, S. Khalifa, and B. W. Schuller, “Domain adapting deep reinforcement learning for real-world speech emotion recognition,” *IEEE Access*, 2024.

[106] M. M. Amin, R. Mao, E. Cambria, and B. W. Schuller, “A wide evaluation of chatgpt on affective computing tasks,” *IEEE Transactions on Affective Computing*, vol. 15, no. 4, pp. 2204–2212, 2024.

[107] D. Melhart, M. Barthet, and G. N. Yannakakis, “Can large language models capture video game engagement?,” *arXiv preprint arXiv:2502.04379*, 2025.

[108] D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, Q. Zhu, S. Ma, P. Wang, X. Bi, *et al.*, “Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning,” *arXiv preprint arXiv:2501.12948*, 2025.

[109] G. Alhussein, I. Ziogas, S. Saleem, and L. J. Hadjileontiadis, “Speech emotion recognition in conversations using artificial intelligence: a systematic review and meta-analysis,” *Artificial Intelligence Review*, vol. 58, no. 7, p. 198, 2025.

[110] X. Rong, “word2vec parameter learning explained,” *arXiv preprint arXiv:1411.2738*, 2014.

[111] B. Patra, V. Suryanarayanan, C. Fufa, P. Bhattacharya, and C. C. Lee, “Scopeit: Scoping task relevant sentences in documents,” in *Proceedings of the 28th International Conference on Computational Linguistics: Industry Track*, pp. 214–227, 2020.

[112] B. Logan *et al.*, “Mel frequency cepstral coefficients for music modeling.,” in *Ismir*, vol. 270, pp. 1–11, Plymouth, MA, 2000.

[113] D. Mehta, M. F. H. Siddiqui, and A. Y. Javaid, “Facial emotion recognition: A survey and real-world user experiences in mixed reality,” *Sensors*, vol. 18, no. 2, p. 416, 2018.

[114] E. Dutta, A. Bothra, T. Chaspari, T. Ioerger, and B. J. Mortazavi, “Reinforcement learning using eeg signals for therapeutic use of

music in emotion management," in *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, pp. 5553–5556, 2020.

[115] N. Deepa, Z. Alkhafajy, S. Kamatchi, *et al.*, "Deep q-network based multi-agent reinforcement learning for driver emotion recognition," in *Proceedings of the International Conference on Networks, Multimedia and Information Technology (NMITCON)*, IEEE, 2025.

[116] M. Barthet, M. Kaselimi, K. Pinitas, K. Makantasis, A. Liapis, and G. N. Yannakakis, "Gamevibe: A multimodal affective game corpus," *Scientific Data*, vol. 11, no. 1306, 2024.

[117] A. Rashed, S. Shirmohammadi, and M. Hefeeda, "Descriptor: Multimodal dataset for player engagement analysis in video games (multipeng)," *IEEE Data Descriptions*, vol. 2, pp. 17–25, 2025.

[118] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. Aerts, "Artificial intelligence in radiology," *Nature Reviews Cancer*, vol. 18, no. 8, pp. 500–510, 2018.

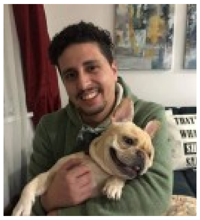
[119] E. De Momi and G. Ferrigno, "Robotic and artificial intelligence for keyhole neurosurgery: the robocast project, a multi-modal autonomous path planner," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 224, no. 5, pp. 715–727, 2010.

[120] S. Borna, M. J. Maniaci, C. R. Haider, C. A. Gomez-Cabello, S. M. Pressman, S. A. Haider, B. M. Demaerschalk, J. B. Cowart, and A. J. Forte, "Artificial intelligence support for informal patient caregivers: a systematic review," *Bioengineering*, vol. 11, no. 5, p. 483, 2024.

[121] Y. Li, M. Wang, L. Wang, Y. Cao, Y. Liu, Y. Zhao, R. Yuan, M. Yang, S. Lu, Z. Sun, *et al.*, "Advances in the application of ai robots in critical care: scoping review," *Journal of medical Internet research*, vol. 26, p. e54095, 2024.

[122] N. Ghotbi, "The ethics of emotional artificial intelligence: a mixed method analysis," *Asian Bioethics Review*, vol. 15, no. 4, pp. 417–430, 2023.

Matthew Barthet received a Bsc. degree in computer science, and a Msc. degree in digital games from the University of Malta in 2019 and 2021, respectively. He is currently a PhD candidate at the University of Malta researching training reinforcement learning agents in affective computing applications. His other research interests include procedural content generation, game artificial intelligence, and computational creativity.



Ahmed Khalifa is an AI researcher/lecturer at the Institute of Digital Games, University of Malta. They got their PhD from New York University back in 2020. They have published more than 60 papers in workshops, conferences, and journals. Their work focused on exploring different methods and techniques for generating game content. They are known for their research on PCGRL, PCG with Quality Diversity, and Deep Tingle. They are also an independent game designer/developer with more than 40 released games/prototypes. Some of their games were nominated for multiple awards in different conferences, such as IndiePrize, Melbourne Queer Games Festival, Queer Games Conference, and International Mobile Game Awards. The games range across different genres such as point-click adventure (Queen Boat), arcade (Atomic+), metroidvania (Hollow Floor), puzzle games (Steps), and word games (worDefense).

Antonios Liapis is an Associate Professor at the Institute of Digital Games, University of Malta, where he bridges the gap between game technology and game design in courses focusing on human-computer creativity, digital prototyping and game development. He received the Ph.D. degree in Information Technology from the IT University of Copenhagen in 2014. His research focuses on Artificial Intelligence in games, human-computer interaction, computational creativity, and user modeling. He has published over 150 papers in the aforementioned fields, and has received several awards for his research contributions and reviewing effort. He serves as Associate Editor for the IEEE Transactions on Games, and has served as general chair in four international conferences, as guest editor in five special issues in international journals, and has co-organized 15 workshops.

Georgios N. Yannakakis is a Professor at the Institute of Digital Games, University of Malta (UM) and a co-founder of humanfeedback.ai and modl.ai. He received the PhD degree in Informatics from the University of Edinburgh in 2006. He does research at the crossroads of artificial intelligence, affective computing, games and computational creativity. He has published more than 400 papers in the aforementioned fields and his work has been cited broadly. His research has been supported by numerous national and European grants (including a Marie Skłodowska-Curie Fellowship) and has appeared in *Science Magazine* and *New Scientist* among other venues. He is currently the Editor in Chief of the *IEEE Transactions on Games*, and used to be Associate Editor of the *IEEE Transactions on Evolutionary Computation*, the *IEEE Transactions on Affective Computing* and the *IEEE Transactions on Computational Intelligence and AI in Games* journals. He has been the General Chair of key conferences in the area of game artificial intelligence (IEEE CIG 2010) and games research (FDG 2013, 2020). Among the several rewards he has received for his papers he is the recipient of the *IEEE Transactions on Affective Computing Most Influential Paper Award* and the *IEEE Transactions on Games Outstanding Paper Award*. Georgios is an IEEE Fellow.