10 Reinforcement Learning in the Realm of Embodied Emotion

Nataša Kovač, Faculty of Applied Sciences, University of Donja Gorica, Oktoih 1, Podgorica, 81000, Montenegro, https://orcid.org/ [0000-0002-6671-2938](https://orcid.org/0000-0002-6671-2938)

Hojjatollah Farahani, Department of Psychology, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran, <https://orcid.org/0000-0002-9799-7008>

Peter Watson, Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge, CB2 7EF, United Kingdom, https://orcid.org/0000-0002-9436-0693

**Abstract:** The chapter explores the groundbreaking intersection of reinforcement learning and emotions, heralding a new frontier in human-computer interaction. The chapter initiates an exploration of emotion-driven reinforcement learning, revealing how emotions can be harnessed to enhance interactions between humans and machines. It then delves into the crucial application of reinforcement learning in monitoring mental health, demonstrating how AI can play a pivotal role in supporting emotional well-being. Additionally, the chapter delves into emotion-driven education, showcasing how reinforcement learning can be tailored to create personalised and emotionally responsive educational experiences. This exploration not only highlights the transformative potential of emotions in AI but also underscores their significance in various facets of our daily lives.

**Keywords:** Emotion-Driven Reinforcement Learning; Monitoring Mental Health; Emotion-Driven Education

# 10.1 Introduction

For decades, psychologists, neuroscientists, and artificial intelligence researchers have sought to decode human emotions, not merely as abstract psychological states but as adaptive responses deeply embedded within dynamic interactions between humans and their environments. Emotions influence behaviour, shape decisions, and significantly affect learning processes through a complex interplay of reward systems, motivation, and behavioural adaptation (Rolls, 2014). However, when emotions are scrutinised through the joint lens of psychological science and computational learning, a far more uncomfortable truth is revealed: many prevailing theories underestimate the ruthlessness of emotional dynamics. Human emotion is not a gentle signalling system; it is an evolved mechanism engineered by natural selection to bias cognition in favour of survival and resource acquisition. Behavioural economics has repeatedly demonstrated that emotional biases systematically distort human judgment, often in predictable, maladaptive ways that deviate from rational-choice principles (Kahneman & Tversky, 1979). From loss aversion to threat amplification, emotional systems operate as diagnostic shortcuts, fast, efficient, and frequently brutal in their overriding of deliberative thought.

In recent years, reinforcement learning (RL), a robust computational paradigm inspired by how biological agents learn from rewards and punishments, has emerged as a persuasive framework for modelling, analysing, and ultimately improving human emotional interactions and experiences (Sutton & Barto, 2018).

RL-based models reinforce this insight, indicating that negative reinforcement, reward prediction errors, and unexpected outcomes produce disproportionately large shifts in behavioural policy updates, mirroring the human tendency to weight aversive information more heavily than positive feedback (Schultz, 2016). At the neural level, emotions are expressed through continuous prediction, evaluation, and error-correction loops that are functionally indistinguishable from RL mechanisms. The dopaminergic system, critical for signalling reward prediction errors, does not merely support learning but also shapes the intensity, duration, and direction of emotional responses (Schultz, 2017). Emotional valence itself can be conceptualised as a dynamic estimate of expected value relative to shifting internal and external contexts. Psychological research demonstrates that chronic stress, trauma, and affect dysregulation distort value-estimation processes, resulting in maladaptive behavioural policies that persist long after environmental contingencies have changed (Pessoa, 2013). In essence, emotional disorders can be understood as pathological reinforcement schedules, overlearning of threat, underlearning of safety, and dysfunctional reward prediction mechanisms that confine individuals to rigid behavioural loops.

Contemporary affective neuroscience has increasingly abandoned the assumption that emotions are static, modular entities. Instead, emotional states are seen as emerging from distributed, context-sensitive appraisal processes integrating sensory cues, bodily states, memory, and social inference (Barrett, 2017). When RL models are applied to such multidimensional psychological processes, they expose a stark mismatch between how humans assume emotion’s function and how they in fact operate. The emotional system is biased, noisy, and strategically tuned to optimise survival rather than well-being. It prioritises rapid threat detection over accuracy, social conformity over autonomy, and energy conservation over psychological flourishing. These computational and psychological insights necessitate a critical reassessment of long-held assumptions in clinical psychology, educational psychology, and behavioural intervention design.

If emotional learning mirrors RL mechanisms, therapeutic change must involve genuine policy updating, not gentle persuasion nor reflective insight, but the systematic restructuring of reward contingencies, exposure patterns, and decision landscapes. Indeed, some of the most effective psychological interventions already rely implicitly on RL principles. Cognitive‑behavioural exposure therapies modify avoidance policies through controlled prediction‑error signalling; behavioural activation reshapes reward schedules to counter depressive withdrawal, and dialectical behaviour therapy manipulates contingencies to stabilise emotional responses. Such approaches function not because they align with comforting clinical narratives but because they force the emotional system to relearn through experience. RL therefore clarifies why human emotion is stubborn, why insight alone rarely transforms behaviour, and why deeply ingrained patterns of fear, attachment, or avoidance resist change until the underlying reinforcement architecture is dismantled and rebuilt. No romanticised or filtered interpretation of human emotional life withstands this computational scrutiny: emotions operate as learning systems, and learning systems are indifferent to human ideals of rationality, morality, or coherence. They follow the mathematics.

Unlike traditional supervised or unsupervised learning approaches, RL explicitly incorporates a temporal dimension of experience. Agents learn sequentially through trial-and-error interactions, adapting continuously based on received feedback signals. This interaction-based learning closely resembles how human emotional systems shape adaptive behaviours over time, often unconsciously guiding our decisions based on perceived emotional outcomes and expectations of future emotional rewards or costs (Damasio, 1996).

Emotional interaction can be framed as an RL control problem with explicit temporal credit assignment: an agent chooses actions, observes their consequences, and updates its policy to maximise long-run return (Sutton & Barto, 2018). To make this safe and useful in human-facing contexts, the learning loop is coupled to evaluation mechanisms and governance constraints. Figure 1 depicts this arrangement: experience is generated online, stored, and then used by a learner/optimiser to update the policy, while value/critic estimates stabilise learning and off-policy evaluation gates deployment. This separation is often termed *offline* or *batch* RL. It has become central when data are sensitive or interventions must not degrade user well-being (Levine et al., 2020).

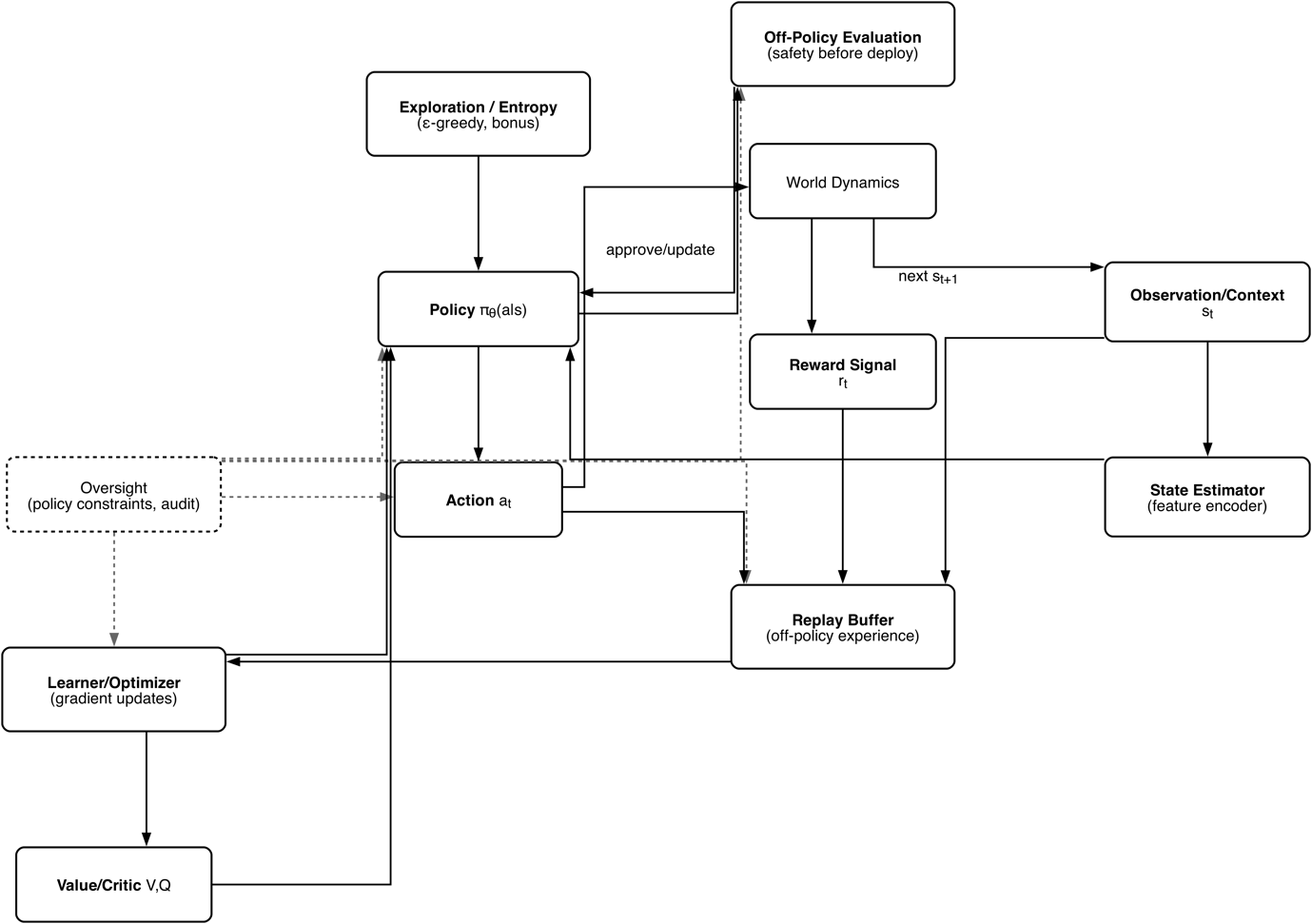


Figure 1. Core reinforcement learning loop with governance constraints and off-policy evaluation

Observations st​ are encoded by a state estimator and passed to the policy πθ​(a∣s), which selects an action at​. The environment (here, the human–computer setting with its world dynamics) returns a reward rt​ and the next state st+1​. Every interaction tuple is logged in a replay buffer, forming a dataset for learning. On the far right, the learner/optimiser performs gradient updates using this logged experience (off-policy), and the stacked value/critic module estimates future returns to reduce variance and improve stability. An exploration/entropy mechanism regularises the policy so it does not collapse prematurely on narrow behaviours, while off-policy evaluation tests candidate policies against the logged data before any change is approved for deployment. Dotted lines indicate governance and safety constraints that *constrain* optimisation rather than being added post hoc (Achiam et al., 2017; Thomas & Brunskill, 2016). Practically, this architecture lets us (i) gather interaction data conservatively, (ii) improve policies offline under explicit constraints, and (iii) deploy only after high-confidence evaluation. This approach is particularly important for emotion-sensitive applications.

By connecting RL with human emotion, we open a new topic for exploration, one that allows intelligent systems not just to recognise emotional states but to actively respond to, learn from, and even influence emotional experiences in context-aware, personalised ways. The state-of-the-art landscape of emotion-informed RL reveals a promising yet emerging field. Early efforts often viewed emotions simply as additional signals to enhance computational decision-making. However, recent advances have marked a paradigm shift, recognising emotions as essential states that influence entire decision processes within interactive settings. Cutting-edge algorithms now utilise emotional signals to adapt the behaviour of intelligent agents in real time. Related machine-learning frameworks have also been applied to affect-linked constructs beyond core emotions, such as love addiction, where feature patterns and explanations clarify predictive factors (Farahani et al., 2025). These agents develop policies that optimise not only traditional task performance but also emotional well-being, engagement, mental health outcomes, and educational effectiveness (Paiva et al., 2017).

A growing body of research highlights that integrating emotional signals into RL significantly enhances the human-computer interaction experience. Emotionally intelligent RL agents have begun to transform key areas such as mental health monitoring, personalised education, assistive robotics, and adaptive human-machine communication. For instance, in mental health, RL models adjust therapeutic interactions in real-time based on subtle emotional feedback, leading to more effective and empathetic mental health interventions (Jaques et al., 2016). In education, personalised learning agents harness emotional feedback to adapt difficulty, pacing, and motivational strategies, significantly improving learner engagement and educational outcomes (D'Mello & Graesser, 2013).

Despite these promising advancements, several challenges remain at the frontier of RL and emotional computing. Complexities inherent to emotional states pose critical questions for computational modelling. RL algorithms must handle noisy, sparse, delayed, and occasionally contradictory emotional feedback. Emotional adaptation through RL must align with stringent ethical guidelines, protecting the privacy, autonomy, and psychological safety of human participants. Thus, the field stands at a compelling intersection, balancing technological sophistication with careful consideration of ethical responsibility and psychological validity.

This chapter embarks on a profound exploration into this emerging landscape, aiming to provide readers with in-depth insights into how RL can effectively model and leverage human emotional dynamics. Through carefully selected case studies, recent developments, critical theoretical discussions, and thoughtful ethical considerations, we delve into how emotion-driven RL is redefining the relationships between humans and intelligent systems, shaping the future of interactive technology and enhancing our understanding of human emotional life.

# 10.2 Emotion-Driven Reinforcement Learning: A New Frontier in Human-Computer Interaction

In recent years, human-computer interaction (HCI) has experienced a significant paradigm shift, moving beyond task-driven exchanges to emotional and socially aware interactions. This integration aligns with the ‘Artificial Psychology’ (PsAIchology) framing that connects AI methods with psychological science (Farahani et al., 2024). Contemporary systems not only respond to explicit user commands but also actively sense, interpret, and adapt dynamically to the emotional states of users. At the heart of this transformation lies emotion-driven RL. This powerful computational framework enables intelligent agents to learn from emotional signals as intrinsic rewards, guiding adaptive, personalised, and context-sensitive interactions (Jaques et al., 2015).

Traditional RL algorithms operate by optimising numerical reward signals, typically aimed at maximising performance in well-defined tasks. However, such conventional approaches fail to reflect the complexity, nuance, and subjectivity of human emotional experiences. Emotion-driven RL fills this critical gap by integrating affective feedback, such as emotional valence, arousal, or psychological engagement, directly into the RL process. This approach fundamentally transforms how artificial agents interact with humans, ensuring responses and adaptations resonate emotionally, leading to deeper engagement and more meaningful interactions (Gratch & Marsella, 2014). For psychology-focused, Python-based modelling and pipeline patterns that complement these designs, see (Kovač et al., 2024). In practice, we map sensed affect to rewards and keep policies sufficiently exploratory to avoid brittle behaviour, as it is shown bellow:

import numpy as np

from typing import Mapping

def affect\_reward(valence: float, arousal: float, goals: Mapping[str, float]) -> float:

dv = valence - goals.get("valence", 0.0)

da = arousal - goals.get("arousal", 0.0)

towards = - (dv\*\*2 + da\*\*2) # closer is better

saturate = - 0.2 \* (max(0, abs(valence)-1)\*\*2 + max(0, abs(arousal)-1)\*\*2)

return float(towards + saturate)

def exploration\_schedule(t: int, eps0: float = 0.2, eps\_min: float = 0.02, half\_life: int = 50\_000) -> float:

import math

return max(eps\_min, eps0 \* math.exp(-t / half\_life))

This code shows a mapping from affect to scalar reward, including the fact that agents should optimise engagement and well-being, not proxy clicks. The second function shows how to keep the agent curious. It represents an implementation detail behind the *Exploration/Entropy* box in Figure 1.

Emotion-driven RL marks a significant advancement in the realm of interactive technologies because emotions often implicitly guide human decision-making, behaviours, and social interactions. Capturing and incorporating these emotional dynamics into machine learning enhances not only the technical performance of HCI systems but also their psychological appropriateness and emotional intelligence, leading to interactions that feel more natural, empathetic, and responsive (Schuller & Schuller, 2018). At its core, emotion-driven RL treats emotional signals as feedback mechanisms or intrinsic rewards. For instance, an intelligent tutoring system can interpret students' emotional states, such as frustration, boredom, engagement, or delight, and adjust its instructional strategies in real-time. Similarly, assistive robots can sense user stress or discomfort, modifying their behaviour proactively to foster comfort and trust (Broekens, Jacobs, & Jonker, 2015).

These emotional reward signals may originate from multiple modalities, including physiological sensors (e.g., heart rate variability, electroencephalogram), facial expression recognition, voice analysis, or textual sentiment analysis. By interpreting these rich emotional inputs, RL algorithms learn policies that are explicitly shaped by human affective responses rather than solely task-based outcomes. Thus, emotion-driven RL agents effectively bridge the gap between computational effectiveness and human psychological needs, preferences, and well-being. Emotion-driven RL has demonstrated substantial promise across diverse interactive domains, including:

* **Affective Computing and Virtual Agents:** Virtual companions and conversational agents now integrate emotional feedback to enhance realism, empathy, and user bonding, leading to higher satisfaction and sustained user engagement (McDuff & Czerwinski, 2018).
* **Adaptive Gaming Experiences:** Games employ real-time emotional adaptation, dynamically adjusting difficulty, narratives, or interactive elements based on detected emotional states, thereby maximising enjoyment and engagement (Yannakakis & Togelius, 2011).
* **Robotics and Socially Assistive Robots:** Robots in caregiving contexts or therapeutic interventions increasingly utilise emotional reinforcement signals to provide personalised emotional support, ensuring interactions remain sensitive, empathetic, and psychologically supportive (Breazeal et al., 2016).

Despite its transformative potential, emotion-driven RL faces unique methodological and practical challenges. Emotional signals are often noisy, ambiguous, context-dependent, and temporally dynamic, complicating their integration into RL frameworks. Key methodological issues include:

* **Temporal Credit Assignment:** Emotional responses can be delayed or ambiguous, complicating accurate attribution of emotional feedback to specific agent actions.
* **Individual Differences and Adaptation:** Emotional patterns and reactions vary significantly among individuals, requiring agents to learn rapidly and generalize robustly across diverse emotional profiles.
* **Ethical and Privacy Concerns:** Leveraging emotional signals necessitates stringent safeguards to respect user privacy, psychological safety, informed consent, and emotional autonomy.

These challenges are balanced by considerable opportunities: emotion-driven RL opens unprecedented possibilities to humanize technology, deepen user engagement, foster emotional well-being, and provide personalized support.

As RL moves deeply into emotional interactions, ethical responsibilities multiply. Systems must not exploit or manipulate emotions, nor compromise user autonomy and privacy. Ethical principles must include:

* **Transparency:** Users should be explicitly informed of how emotional data is collected, interpreted, and utilized.
* **Consent and Autonomy:** Emotional interactions must respect individual autonomy, offering clear mechanisms for users to manage or withdraw emotional data use.
* **Psychological Safety:** Emotion-driven interactions must prioritize emotional safety, ensuring interventions remain supportive rather than intrusive or coercive.

Figure 2 facilitates a deeper conceptual understanding of RL. This integrated conceptual representation illustrates the sophisticated interaction among emotional sensing, RL agent adaptation, user interaction, and ethical oversight.

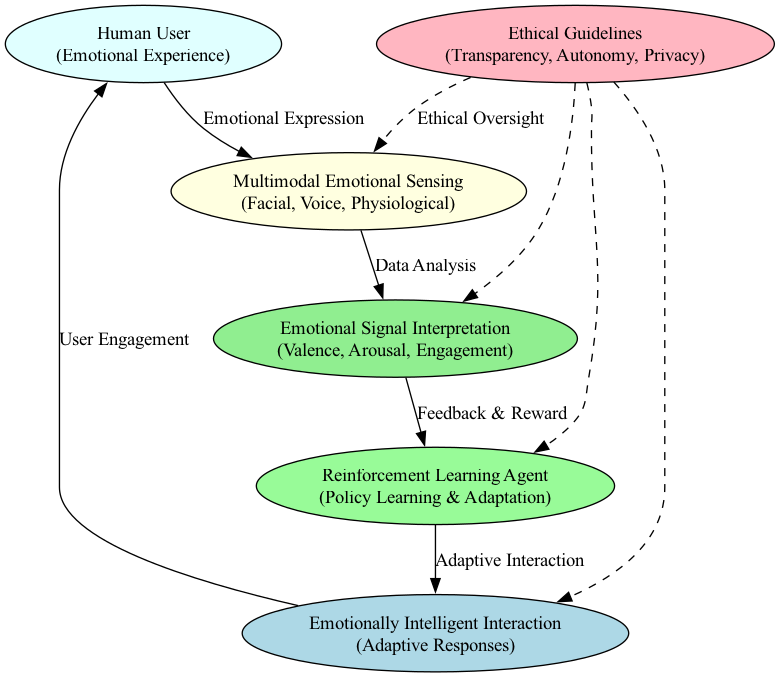


Figure 2. Emotion-driven reinforcement learning framework

This conceptual framework elucidates the complex, dynamic interactions in emotion-driven RL systems, highlighting how emotional sensing and ethical oversight comprehensively shape interactive and adaptive processes. Solid arrows show the fast control loop from user expression to sensing, analysis, interpretation, RL adaptation, and back to the user. Dashed arrows show cross-cutting ethical oversight that constrains data capture, feature use, reward design, and allowable system actions, ensuring transparency, autonomy, and privacy throughout the loop.

# 10.3 Monitoring Mental Health with Reinforcement Learning

Mental health represents one of society's most pressing challenges, profoundly impacting quality of life, productivity, and well-being globally. Traditional approaches to monitoring and treating mental health issues rely largely on periodic assessments and self-reported measures, typically captured through clinical consultations, structured questionnaires, or interviews. However, these methods often lack granularity, timeliness, and ecological validity, failing to capture subtle day-to-day emotional fluctuations crucial for proactive intervention (Mohr, Zhang, & Schueller, 2017).

The emergence of RL as a powerful computational paradigm in healthcare and mental health management has introduced transformative possibilities, shifting from reactive or periodic care towards proactive, continuous, and contextually sensitive mental health monitoring and support (Luo et al., 2022). RL methods, particularly when informed by embodied emotional signals such as physiological data, speech patterns, facial expressions, and behavioural cues, enable intelligent systems to dynamically adapt interactions, interventions, and monitoring strategies in real-time based on the user's mental and emotional states (Jaques et al., 2017). Comparable supervised pipelines are also used across psychological health domains, for example the classification of chronic pain outcomes (Kovač et al., 2025). Relatedly, regression pipelines have been used to predict internal shame from psychological predictors, demonstrating explainable ML in affect-linked constructs (Kovač, Ratković, Farahani, & Watson, 2025a).

In conventional mental health monitoring, assessment intervals often span weeks or even months, limiting responsiveness and adaptive intervention. In contrast, RL agents continuously interact with individuals, learning from subtle emotional and behavioural signals, allowing early detection of emotional distress or mental health deterioration, and proactively adapting interventions to support mental health more effectively. This continuous adaptation approach significantly improves treatment personalisation and timeliness, directly addressing fluctuations in mental health conditions such as depression, anxiety, and stress disorders (Alasmrai et al., 2025).

RL-based monitoring systems leverage real-time multimodal emotional data (including speech analysis, wearable sensor data, facial emotion analysis, and digital behavioural patterns) to dynamically adjust their actions. These actions may involve alerting human caregivers, modifying therapeutic exercises, initiating supportive conversations through virtual agents, or recommending personalised self-care strategies. Importantly, the adaptability inherent to RL allows these interventions to evolve continuously, maximising emotional support and psychological benefit while minimising intrusiveness or discomfort.

Several pioneering applications of RL for mental health monitoring highlight the powerful potential of this approach:

* **Adaptive Digital Interventions**: RL systems have been integrated into smartphone-based mental health apps, dynamically adjusting intervention types and frequencies based on real-time assessments of emotional well-being (Liao et al., 2020).
* **Intelligent Conversational Agents and Chatbots**: Virtual mental health assistants utilize RL algorithms to adapt their conversational style, therapeutic strategies, and content dynamically, effectively responding to user emotional states and promoting emotional comfort and engagement (Fitzpatrick, Darcy, & Vierhile, 2017).
* **Wearable-Based Stress and Anxiety Management**: Wearable devices combined with RL-driven apps interpret continuous physiological data, such as heart rate variability and skin conductance, to personalize stress-reduction interventions in real-time, helping users maintain emotional balance and psychological well-being proactively (Rohani et al., 2018).

Implementing RL in mental health monitoring involves specific methodological considerations unique to this domain:

* **Sparse and Ambiguous Emotional Rewards**: Emotional and mental health signals are inherently complex, noisy, and ambiguous, making accurate reward interpretation challenging for RL agents.
* **Delayed and Long-Term Outcomes**: Emotional and mental health improvements often manifest gradually, posing significant temporal credit-assignment challenges for RL algorithms.
* **Personalization and Generalization**: Mental health states vary dramatically among individuals, necessitating RL models capable of personalized adaptation and rapid generalization across diverse user profiles.

These challenges, however, provide significant opportunities for methodological innovation, such as sophisticated hierarchical RL architectures, contextual bandit algorithms, and advanced reward modelling techniques tailored specifically for mental health contexts.

RL applications in mental health raise critical ethical questions. Mental health data, by its very nature, is extremely sensitive, demanding strict adherence to ethical principles, including:

* **Privacy and Data Security**: Continuous emotional monitoring must rigorously ensure user confidentiality, informed consent, and secure data handling, preventing unauthorised access or unintended use.
* **Transparency and Explainability**: Users should clearly understand how their emotional data is collected, interpreted, and used for RL-driven adaptations, maintaining transparency and trust.
* **Psychological Safety and Autonomy**: RL-driven interventions should respect user autonomy, ensuring that proactive support remains non-intrusive, emotionally safe, and genuinely supportive rather than coercive or manipulative.

Those three principles aren’t an afterthought. They are shaping the whole system. A rigorous design treats ethics as architectural constraints, not post-hoc add-ons. Privacy and security, transparency and explainability, and psychological safety and autonomy are specified a priori and enforced across sensing, inference, learning, and action. Figure 3 operationalises this approach by placing a governance layer that constrains data capture, feature use, policy optimisation, and escalation pathways throughout the system.

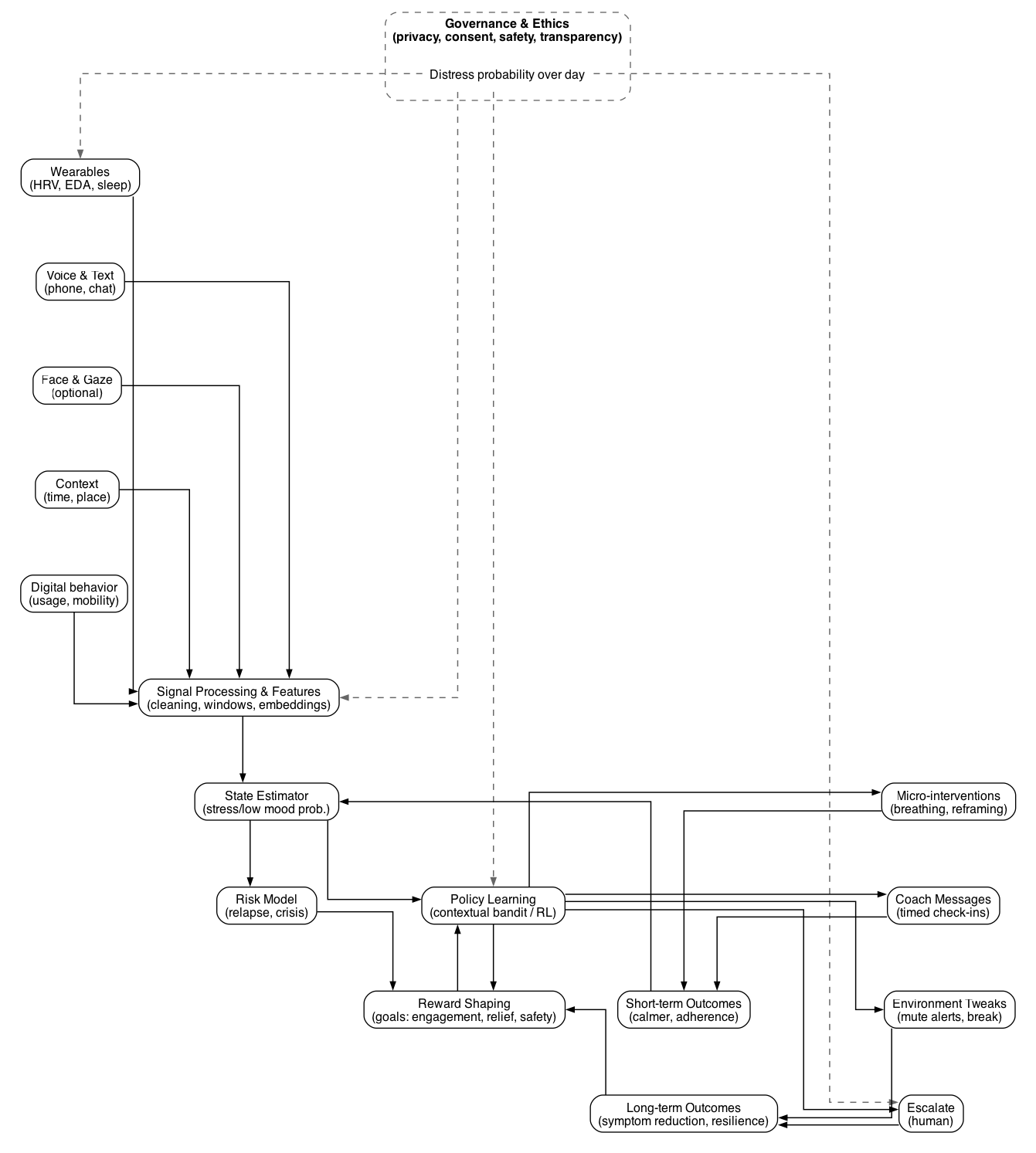


Figure 3. Monitoring mental health with reinforcement learning

The dashed band across the top (Governance & Ethics) is a reminder that every layer beneath operates under explicit constraints: what can be sensed (consent), how it is stored (privacy and security), how inferences are reported (transparency and explainability), and what actions are allowed (psychological safety and user autonomy). The dashed connectors are not data pipes; they are *oversight links* showing that governance continuously limits and guides sensing, feature use, learning, and escalation.

Beneath that, the signals row shows the kinds of input we might collect with consent: wearables (HRV, EDA, sleep), voice and text, optional face and gaze, simple context (time/place), and digital behaviour (e.g., phone-use rhythms). These raw streams are not yet emotions. They flow into Signal Processing & Features, where they’re cleaned, synchronised, and summarised without discarding affect-laden cues. From there, two kinds of inferences are made. A State Estimator produces moment-to-moment probabilities (e.g., *low mood likely*), while a Risk Model looks ahead to estimate relapse or crisis risk. Keeping state and risk separate helps with explainability: one tells you *how things look now*, the other *what seems likely next*.

The following functions show a minimal contextual bandit for micro-interventions, risk-aware shaping, and a safe escalation rule that mirrors the graded actions in Figure 3:

import numpy as np

def softmax(x: np.ndarray, temp: float = 0.5) -> np.ndarray:

z = x / max(1e-6, temp)

z -= z.max()

e = np.exp(z)

return e / e.sum()

def choose\_intervention(context\_phi: np.ndarray, W: np.ndarray, temp: float = 0.5) -> int:

scores = W @ context\_phi

probs = softmax(scores, temp=temp)

return int(np.random.choice(len(scores), p=probs))

def risk\_shaped\_reward(relief: float, engagement: float, risk\_prob: float,

w\_relief=0.6, w\_eng=0.4, risk\_penalty=2.0) -> float:

base = w\_relief \* relief + w\_eng \* engagement

return float(base - risk\_penalty \* max(0.0, risk\_prob))

from dataclasses import dataclass

@dataclass

class EscalationPolicy:

hi: float = 0.7 # high risk threshold

lo: float = 0.4 # de-escalation threshold

min\_cooldown: int = 3 # interactions between escalations

def should\_escalate(risk\_now: float, last\_escalation\_steps\_ago: int, policy: EscalationPolicy) -> bool:

return (risk\_now >= policy.hi) and (last\_escalation\_steps\_ago >= policy.min\_cooldown)

Decisions live in the next row. Policy learning (the RL component) selects actions, but it does so alongside reward shaping, where goals are set in human terms (*relief, engagement, safety*), not raw screen time. That’s the practical expression of autonomy and safety from the bullet list above: the agent learns, but only within guardrails that favour well-being. Actions are graded. The system starts with micro-interventions (a breathing cue, a brief reframing), then coach messages (timed, light-touch check-ins), then environment tweaks (quieting notifications or suggesting a break). Only when risk warrants it does the system escalate to a human. This tiered design keeps help supportive, not intrusive, and it leaves room for human judgment when it matters most. Short-term (calmer now? adhered?) and long-term (symptom reduction, resilience) outcomes flow back to update the estimator and the learning policy. That feedback loop is how the system becomes more personal over time *without* relaxing its ethical boundaries.

# 10.4 Emotion-Driven Education

Education has traditionally emphasised cognitive dimensions, such as knowledge acquisition, logical reasoning, and skill development. Yet, decades of educational and psychological research underscore that learning is fundamentally intertwined with emotion. Emotional states, such as curiosity, frustration, boredom, and enjoyment, critically influence learners' motivation, engagement, retention, and overall academic achievement (Pekrun & Linnenbrink-Garcia, 2014). Despite this understanding, conventional educational systems often lack the flexibility and responsiveness required to address dynamic emotional experiences, relying instead on static instructional designs and uniform learning strategies.

Emotion-driven education, empowered by RL, represents a transformative shift toward responsive, personalised, and emotionally sensitive educational experiences. RL methodologies enable educational systems to continuously monitor learners' emotional states and dynamically adapt educational content, interactions, and strategies accordingly, optimising both learning outcomes and emotional well-being (D’Mello & Graesser, 2012).

At the heart of emotion-driven education lies the recognition that emotions are not mere secondary variables in the learning process but central drivers of educational success. RL uniquely addresses this perspective by treating emotional signals as key feedback mechanisms guiding instructional adaptation. Unlike traditional rule-based or predetermined interventions, RL-driven educational systems learn continuously from students' emotional responses, adapting in real-time to individual emotional profiles, preferences, and needs (Calvo & D'Mello, 2010). Figure 4 visualises a simple emotion-aware *policy surface*. It shows how an RL tutor can map combinations of learner arousal and task challenge to pedagogical actions, and how a learner’s state moves through this space during a session (Pekrun, 2018; Baker et al., 2010).

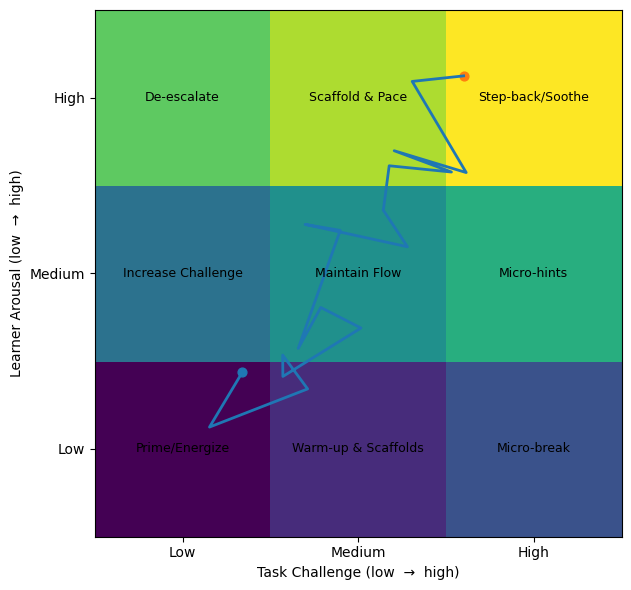


Figure 4. Emotion-aware policy surface with session trajectory

Rows represent learner arousal (low–high) and columns task challenge (low–high). Each cell names the *recommended action* (e.g., *Maintain Flow*, *Micro-break*, *Scaffold & Pace*). The polyline traces a synthetic session: the learner starts under-engaged, enters a productive *flow* region (moderate arousal, moderate challenge), then spikes toward overload, prompting actions that de-escalate and re-scaffold. The goal is not to push difficulty relentlessly, but to stabilise around flow while preserving motivation and affective safety (Pekrun, 2018). Seen this way, affect becomes part of the state the tutor optimises: low arousal + low challenge calls for energising prompts; medium + medium favours holding pace; high arousal + high challenge triggers step-back or soothing actions before learning resumes. Empirically, keeping learners near that middle band is associated with better persistence, deeper processing, and fewer unproductive episodes of boredom or frustration (Baker et al., 2010). The policy surface is thus a compact picture of *how* an emotion-sensitive RL system operationalises the theories you cite. The next function encodes the policy surface from Figure 4, and a small utility turns sensed arousal–challenge into an action label:

from typing import Literal

Action = Literal["Energise", "Maintain Flow", "Micro-break", "Scaffold & Pace", "De-escalate"]

def tutor\_policy\_surface(arousal: float, challenge: float) -> Action:

a, c = float(arousal), float(challenge)

if a < 0.33 and c < 0.33: # bored & easy

return "Energise"

if 0.33 <= a <= 0.66 and 0.33 <= c <= 0.66:

return "Maintain Flow"

if a > 0.66 and c > 0.66: # overload

return "De-escalate"

if a > 0.66 and c <= 0.66: # anxious, moderate task

return "Micro-break"

return "Scaffold & Pace" # low arousal, high challenge or mixed

Emotion-driven educational systems employ multimodal emotional sensing methods to detect learners' emotional states through facial expressions, gaze patterns, physiological responses (e.g., heart rate, skin conductance), voice dynamics, and interaction patterns (e.g., mouse movements, task completion times). By interpreting these emotional signals as intrinsic rewards or penalties, RL algorithms dynamically adjust learning difficulty, pacing, instructional style, content presentation, and motivational strategies to sustain engagement and support positive emotional experiences (Baker et al., 2010).

RL has already shown notable potential in emotion-driven education, with diverse applications illustrating its transformative possibilities:

* **Intelligent Tutoring Systems (ITS)**: Advanced ITS integrate RL-based emotion sensing to adjust instructional interventions dynamically, recognising frustration or disengagement and responding with personalised scaffolding, hints, or encouragement, thereby maintaining learner motivation and engagement (Arroyo et al., 2011).
* **Adaptive Educational Games**: Educational gaming environments leveraging RL can identify emotional states in real-time, adapting gameplay difficulty and providing emotional support through tailored feedback, maintaining optimal challenge levels, and maximising enjoyment and learning efficacy (Conati & Maclaren, 2009).
* **Virtual Learning Environments (VLEs)**: Emotion-aware VLEs utilise RL to proactively manage learner emotions, adjusting content and interactions dynamically to reduce stress and anxiety, thereby improving overall learning experiences and outcomes, especially in remote or virtual educational contexts (Shute & Ke, 2012).

Integrating RL into emotion-driven educational systems presents several methodological challenges and opportunities:

* **Complexity of Emotional Signals**: Accurately interpreting subtle emotional indicators in educational settings requires sophisticated, multimodal data integration techniques to handle noisy, ambiguous, and context-dependent signals.
* **Delayed and Ambiguous Emotional Rewards**: Learner emotions can fluctuate rapidly, often delayed or influenced by external factors, making RL credit assignment particularly challenging.
* **Personalisation at Scale**: Effective RL-driven educational systems must rapidly adapt to diverse emotional profiles across a large student population, balancing individualised adaptation with scalability and generalizability.

These challenges present significant research opportunities, inviting the development of advanced hierarchical RL architectures, robust emotional signal interpretation methods, and innovative personalised learning strategies. Because students change over time, a simple **affect-stabilising curriculum** helps the tutor stay near the flow region while still progressing content. The next function implements the *keep near flow* principle (gentle push when calm, stronger pullback when arousal spikes):

def curriculum\_step(difficulty: float, arousal: float, target\_arousal: float = 0.5,

k\_up: float = 0.05, k\_down: float = 0.08) -> float:

if arousal <= target\_arousal:

difficulty += k\_up

else:

difficulty -= k\_down

return float(min(1.0, max(0.0, difficulty)))

As educational systems delve deeper into emotional monitoring and adaptive interactions, critical ethical responsibilities emerge:

* **Privacy and Data Protection**: Monitoring emotional states requires strict data privacy protocols, ensuring emotional data remains confidential, secure, and used exclusively for educational enhancement.
* **Transparency and Explainability**: Students, parents, and educators must fully understand how emotional data informs adaptive educational strategies, ensuring transparency, consent, and trustworthiness.
* **Autonomy and Emotional Safety**: RL-driven adaptations must carefully preserve student autonomy and emotional safety, avoiding manipulative practices, and always prioritising emotional well-being alongside academic success.

**References:**

Achiam, J., Held, D., Tamar, A., & Abbeel, P. (2017, July). Constrained policy optimization. In *International conference on machine learning* (pp. 22-31). PMLR.

Alasmrai, M. A., Ismail, R. M., & Ali Al-Abyadh, M. H. (2025). Personalized Cognitive Behavioral Therapy for Adults Using Machine Learning: A Multi-Factor, Reinforcement-Based Approach. *Fusion: Practice & Applications*, *20*(2).

Arroyo, I., Woolf, B. P., Cooper, D. G., Burleson, W., & Muldner, K. (2011, July). The impact of animated pedagogical agents on girls' and boys' emotions, attitudes, behaviors and learning. In *2011 IEEE 11th international conference on Advanced learning technologies* (pp. 506-510). IEEE.

Baker, R., D’Mello, S., Rodrigo, M., & Graesser, A. (2010). Better to be frustrated than bored: The incidence and persistence of affect during interactions with three different computer-based learning environments. *International Journal of human-computer studies*, *68*(4), 223-241.

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Pan Macmillan.

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016). Social robotics. *Springer handbook of robotics*, 1935-1972.

Broekens, J., Jacobs, E., & Jonker, C. M. (2015). A reinforcement learning model of joy, distress, hope and fear. *Connection Science*, *27*(3), 215-233.

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, *1*(1), 18-37.

Cañamero, L., & Lewis, M. (2016). Making new “New AI” friends: Designing a social robot for diabetic children from an embodied AI perspective. *International Journal of Social Robotics*, *8*(4), 523-537.

Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, *19*(3), 267-303.

D’Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145-157.

D'Mello, S., & Graesser, A. (2013). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *2*(4), 1-39.

Damasio, A. R. (1996). *El error de Descartes*. Andrés Bello.

Doya, K. (2008). Modulators of decision making. *Nature neuroscience*, *11*(4), 410-416.

Etkin, A., Büchel, C., & Gross, J. J. (2015). The neural bases of emotion regulation. *Nature reviews neuroscience*, *16*(11), 693-700.

Farahani, H., Watson, P., Bezan, T., Kovač, N., Winter, L. C., Blagojević, M., ... & Jiménez, P. (2024). PsAIchology: An Intelligent Direction in Psychological Sciences. In *10th International Scientific Conference Technics, Informatics and Education-TIE 2024*. Faculty of Technical Sciences Čačak, University of Kragujevac.

Farahani, H., Watson, P., Kovač, N., Sheykhangafshe, F. B., Azadfallah, P., Allahyari, A., ... & Chesli, R. R. (2025). Unraveling the Complexity of Love Addiction Using Machine Learning Algorithms: The Influence of Positive and Negative Affect, Interpersonal Needs, and Self-Hate. In *International Handbook of Love: Transcultural and Transdisciplinary Perspectives* (pp. 1-22). Cham: Springer Nature Switzerland.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health*, *4*(2), e7785.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, *11*(2), 127-138.

Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry*, *1*(2), 148-158.

Gratch, J., & Marsella, S. (2014). *Appraisal models* (pp. 54-67). Oxford University Press.

Huys, Q. J., & Dayan, P. (2009). A Bayesian formulation of behavioral control. *Cognition*, *113*(3), 314-328.

Jaques, N., McDuff, D., Kim, Y. L., & Picard, R. (2016, September). Understanding and predicting bonding in conversations using thin slices of facial expressions and body language. In *International Conference on Intelligent Virtual Agents* (pp. 64-74). Cham: Springer International Publishing.

Jaques, N., Taylor, S., Sano, A., & Picard, R. (2015, December). Multi-task, multi-kernel learning for estimating individual wellbeing. In *Proc. NIPS Workshop on Multimodal Machine Learning, Montreal, Quebec* (Vol. 898, p. 3).

Joffily, M., & Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS computational biology*, *9*(6), e1003094.

Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I* (pp. 99-127).

Kahnt, T., & Tobler, P. N. (2016). Dopamine regulates stimulus generalization in the human hippocampus. *Elife*, *5*, e12678.

Koban, L., & Wager, T. D. (2016). Beyond conformity: Social influences on pain reports and physiology. *Emotion*, *16*(1), 24.

Kovač, N., Ratković, K., Farahani, H., & Watson, P. (2024). A practical applications guide to machine learning regression models in psychology with Python. *Methods in Psychology*, *11*, 100156.

Kovač, N., Ratković, K., Farahani, H., & Watson, P. (2025a). Machine learning regression models for internal shame. Acta Psychologica, 260, 105721.

Kovač, N., Ratković, K., Watson, P., Farahani, H., & Bagheri Sheykhangafshe, F. (2025). Machine learning classification models for predicting chronic pain. *Current Psychology*, 1-14.

Levine, S., Kumar, A., Tucker, G., & Fu, J. (2020). Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*.

Liao, P., Greenewald, K., Klasnja, P., & Murphy, S. (2020). Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, *4*(1), 1-22.

Luo, Y., Wunderink, R. G., & Lloyd-Jones, D. (2022). Proactive vs reactive machine learning in health care: lessons from the COVID-19 pandemic. *Jama*, *327*(7), 623-624.

McDuff, D., & Czerwinski, M. (2018). Designing emotionally sentient agents. *Communications of the ACM*, *61*(12), 74-83.

Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annual review of clinical psychology*, *13*(1), 23-47.

Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *7*(3), 1-40.

Pekrun, R. (2018). Control-value theory: A social-cognitive approach to achievement emotions. *Big theories revisited*, *2*, 162-190.

Pekrun, R., & Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. In *International handbook of emotions in education* (pp. 1-10). Routledge.

Pessoa, L. (2013). *The cognitive-emotional brain: From interactions to integration*. MIT press.

Rohani, D. A., Faurholt-Jepsen, M., Kessing, L. V., & Bardram, J. E. (2018). Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review. *JMIR mHealth and uHealth*, *6*(8), e9691.

Rolls, E. T. (2014). *Emotion and decision making explained*. OUP Oxford.

Schuller, D., & Schuller, B. W. (2018). The age of artificial emotional intelligence. *Computer*, *51*(9), 38-46.

Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. *Nature reviews neuroscience*, *17*(3), 183-195.

Schultz, W. (2017). Reward prediction error. *Current Biology*, *27*(10), R369-R371.

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160007.

Shute, V. J., & Ke, F. (2012). Games, learning, and assessment. In *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 43-58). New York, NY: Springer New York.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). Cambridge: MIT press.

Thomas, P., & Brunskill, E. (2016, June). Data-efficient off-policy policy evaluation for reinforcement learning. In *International conference on machine learning* (pp. 2139-2148). PMLR.

Yannakakis, G. N., & Togelius, J. (2011). Experience-driven procedural content generation. *IEEE Transactions on Affective Computing*, *2*(3), 147-161.