The Role of Personalized Reward Mechanisms in Deep Reinforcement Learning Driven Cognitive Training: Applications, Challenges, and Future Directions

Qimiao Gao
Department of Computer Science, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

Keywords: Deep Reinforcement Learning, Personalized Reward Mechanisms, Cognitive Training, Neurofeedback.

Abstract: Deep Reinforcement Learning (DRL) has revolutionized the field of cognitive training by integrating the decision-making capabilities of Reinforcement Learning (RL) and the perceptual power of Deep Learning (DL). A key component of DRL is the use of personalized reward mechanisms, which dynamically adjust the reinforcement signals to optimize individual learning trajectories. This review explores the application of personalized reward strategies, such as Q-learning, Advantage Actor-Critic (A3C), and Proximal Policy Optimization (PPO), in neurofeedback (NF) interventions for cognitive enhancement. We focus on their roles in treating conditions like attention deficit hyperactivity disorder (ADHD) and anxiety disorders and discuss their effectiveness in virtual reality-based cognitive training environments. Personalized reward mechanisms have shown significant potential in improving learning outcomes, engagement, and motivation by tailoring the difficulty and feedback of tasks to the user's physiological and behavioral states. Despite these successes, challenges remain in Electroencephalography (EEG) data's real-time processing and personalized interventions' scalability across diverse populations. Future research should focus on improving the adaptability and generalization of these reward systems through multimodal data integration and advanced DRL techniques, while also addressing ethical concerns related to data privacy and user well-being.

1 INTRODUCTION

Cognitive training has gained substantial interest in recent years due to its potential to enhance cognitive function across different age groups and populations. It is designed to improve specific cognitive abilities, such as attention, memory, and executive function, through systematic practice. In neurofeedback (NF), cognitive training typically employs methods that enable individuals to regulate their brain activity via real-time feedback, thereby improving attention and executive functions (Enriquez-Geppert, Huster, & Herrmann, 2017).

Deep reinforcement learning (DRL) has emerged as a powerful tool for enhancing the efficacy of NF interventions by personalizing the feedback and task difficulty based on the user's brain signals and performance (Mnih et al., 2015). The application of DRL in cognitive training allows for more adaptive, personalized approaches that can better address individual needs, thereby optimizing learning outcomes and cognitive improvement.

A personalized reward mechanism in reinforcement learning (RL) refers to dynamically adjusting the reward system based on an individual's behavior and performance to optimize learning. In cognitive training, such mechanisms are critical as they help maintain participant engagement and adapt the training to individual differences in cognitive function and learning pace (Sutton, & Barto, 2018). This personalized approach has been more effective than fixed reward strategies because it aligns the reinforcement signal with each individual's learning trajectory, improving motivation and outcomes.

This review aims to explore the role of personalized reward mechanisms in DRL-driven cognitive training, focusing on their applications, challenges, and future directions. The review is arranged as follows: The initial section will provide the theoretical basis for personalized reward mechanisms and DRL, as depicted in Figure 1. We will subsequently analyze their applications in NF interventions for attention deficit hyperactivity disorder (ADHD), anxiety disorders, and cognitive

Gao and Q.
The Role of Personalized Reward Mechanisms in Deep Reinforcement Learning Driven Cognitive Training: Applications, Challenges, and Future Directions.
DOI: 10.5220/0013528100004619
In *Proceedings of the 2nd International Conference on Data Analysis and Machine Learning (DAML 2024)*, pages 543-551
ISBN: 978-989-758-754-2

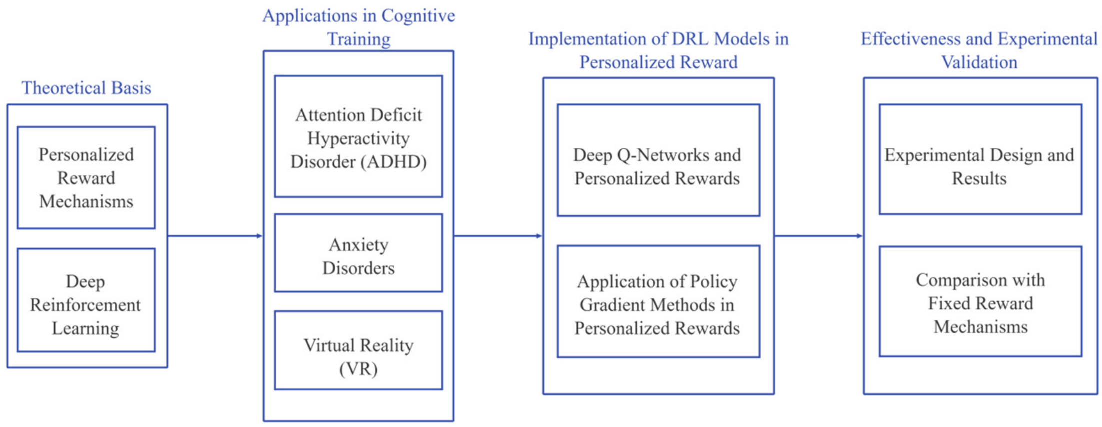

Figure 1: The overall framework of this paper.

training in virtual reality (VR) environments. Following that, we will examine the application of DRL models in personalized reward systems, encompassing Deep Q-Networks (DQN) and policy gradient techniques. Ultimately, we evaluate the effectiveness of these mechanisms, delineate current challenges, and propose recommendations for future research.

2 THEORETICAL BASIS

2.1 Deep Reinforcement Learning

Deep reinforcement learning is a machine learning methodology integrating RL with deep neural networks, allowing agents to derive optimal policies from unprocessed input data (Mnih et al., 2015). The core principle of deep learning (DL) is to employ multi-layered network architectures and nonlinear transformations to integrate low-level features, creating abstract, easily identifiable high-level representations, thus uncovering the distributed feature representations of data. The fundamental concept of RL is to ascertain the optimal policy for attaining a specified objective by maximizing the cumulative reward obtained by the agent through interactions with the environment (Sutton & Barto, 2018). Consequently, DL methods concentrate on the perception and representation of objects, whereas reinforcement learning methods prioritize the acquisition of strategies for problem-solving. Consequently, Google's DeepMind, an AI research division, integrated the perceptual faculties of DL with the decision-making prowess of RL, establishing a novel research focal point in artificial intelligence — DRL. Since then, the DeepMind team has developed and deployed human expert-level agents across numerous challenging domains. These agents construct and acquire knowledge from unprocessed input signals autonomously, without necessitating manual coding or specialized domain expertise. Therefore, DRL is an end-to-end perception and control system with strong generalization capabilities.

The process of learning can be delineated as follows: (1) The agent continuously interacts with the environment, acquires a high-dimensional observation, and employs deep learning techniques to interpret the observation, yielding both abstract and specific state feature representations; (2) The agent assesses the value function of each action predicated on the anticipated return and correlates the current state to the appropriate action via a defined policy; (3) The environment reacts to this action, and the agent obtains the subsequent observation. By continuously iterating through this process, the optimal policy to achieve the goal can be obtained. The theoretical framework of DRL is shown in Figure 2.

In the DRL framework, the agent interacts with an environment, observes states, takes actions, and receives rewards (Sutton & Barto, 2018). The goal is to maximize cumulative rewards by learning an optimal policy that maps states to actions. DRL has been successfully applied to cognitive training to personalize learning experiences and optimize outcomes based on individual user behaviors (Watanabe, Sasaki, Shibata & Kawato, 2018). By modeling complex environments and learning from rich sensory data, DRL provides a powerful tool for adaptive interventions in NF and cognitive enhancement.

Figure 2: DRL Theoretical Framework.

Table 1. Summary of Personalized Reward Strategies in DRL

Strategy	Key Characteristics	Applications
Q-learning-based Reward Adjustment	Adjusts reward function based on individual progress Uses Q-values to estimate action-value pairs	NF interventions for cognitive training ADHD treatment using personalized feedback
Proximal Policy Optimization	Policy gradient method Optimizes training tasks in real time Ensures stable updates to policy	Personalized task difficulty adjustment in cognitive training Electroencephalography (EEG) - based NF for anxiety management
Advantage Actor-Critic	Combines value-based and policy-based methods Multi-agent learning enables rapid adaptation	Real-time personalized feedback in complex environments VR-based cognitive training with EEG signals

2.2 Personalized Rewards

Reward mechanisms play a central role in guiding the behavior of learning agents by providing feedback on the success of actions taken in a given state (Sutton & Barto, 2018). In traditional RL, fixed rewards are used to reinforce desired behaviors, but this approach can be limited when dealing with complex human learning tasks. Personalized rewards, which adapt based on an individual's performance and learning trajectory, have been shown to be significantly more effective in optimizing cognitive outcomes (Silver et al., 2017). Personalized reward mechanisms can maintain learner engagement and motivation, which are crucial for successful cognitive training, especially in NF settings where learning depends heavily on individual differences (D'Esposito, 2008). As shown in Table 1, various personalized reward strategies have been developed to enhance the learning experience in cognitive training. One such strategy is Q-learning-based reward adjustment, where the reward function is tailored to reflect individual progress and specific learning needs (Watkins & Dayan, 1992). Another approach involves policy gradient methods, such as Proximal Policy Optimization (PPO) and Advantage Actor-Critic (A3C), which optimize rewards in real time to maximize the effectiveness of training sessions (Schulman, Wolski, Dhariwal, Radford & Klimov, 2017). These techniques have been successfully applied to NF, providing tailored interventions that dynamically adjust training tasks and reinforcement signals based on individual performance metrics (Watanabe et al., 2018).)

3 APPLICATIONS

3.1 ADHD

Personalized reward mechanisms are particularly beneficial in cognitive training because they provide customized reinforcement based on the individual’s responses and progress, thus optimizing learning outcomes (Sutton & Barto, 2018). In ADHD, individuals often exhibit challenges in maintaining attention and require adaptive strategies to stay engaged in cognitive training sessions (Enriquez-

Geppert et al., 2017). Personalized rewards can be tailored to each individual's learning pattern, which helps in maintaining motivation and ensuring that the interventions are appropriately challenging, yet achievable (Arns et al., 2020). Electroencephalography (EEG)-based NF is well-suited for this personalization, as it provides real-time insights into an individual’s neural activity, enabling adaptive adjustments to the training protocols.

In VR environments, personalized reward strategies can enhance the immersive experience by tailoring the difficulty and feedback based on the user's physiological state and behavior. This personalization not only makes the VR experience more engaging but also promotes better cognitive outcomes by providing optimal challenges suited to each user's cognitive abilities. The combination of VR with EEG signals further enhances the potential for personalized feedback, ensuring that users receive interventions that are responsive to their immediate neural states (Bouchard, Bernier, oivin, Morin & Robillard, 2012).

3.1.1 Application Examples

EEG-based NF combined with DRL has proven effective for personalizing interventions in children and adults with ADHD. In these interventions, EEG signals are used to monitor brain activity, and DRL algorithms adjust NF protocols in real-time to optimize learning outcomes (Enriquez-Geppert et al., 2017; Watanabe et al., 2018). Recent studies have demonstrated that personalized NF using DRL can improve attention and reduce hyperactivity symptoms more effectively compared to conventional methods (Arns et al., 2020). For example, the application of DRL in theta/beta ratio (TBR) NF has shown significant improvements in ADHD patients' cognitive performance, making it a promising treatment alternative to medication (Enriquez-Geppert et al., 2019).

The following provides a comparative overview of EEG-based NF treatments, emphasizing the advantages of personalized interventions for ADHD, as discussed by Garcia Pimenta, Brown, Arns, and Enriquez-Geppert (2021). Personalized reward mechanisms, adapted to each individual's EEG characteristics, have been found to significantly enhance treatment outcomes, yielding a remission rate of 57%, which surpasses that of methylphenidate (31%) and matches the results of medication alone in controlled trials (56%). Techniques such as slow cortical potential (SCP), TBR, and sensorimotor rhythm (SMR) training, when combined with various control conditions and pharmacological treatments, are more effective in improving cognitive performance and reducing ADHD symptoms than conventional methods. Incorporating multimodal strategies, including pharmacotherapy or lifestyle adjustments, further increases the clinical effectiveness of these customized NF interventions.

3.1.2 Real-Time Adaptive Task Generation

Real-time adaptive task generation is crucial in NF interventions to address the specific needs of ADHD patients. DRL can be used to dynamically adjust task difficulty and feedback based on the patient’s real-time EEG data, thereby maintaining optimal engagement and promoting effective learning (Cohen et al., 2015). This personalized approach enables adaptive task settings that align with each patient’s cognitive capacity, ensuring that the challenges are neither too easy nor too difficult. Such personalized adjustments have been found to improve both the effectiveness of the NF training and the motivation of the participants (Sitaram et al., 2016).

3.2 Anxiety Disorders

EEG-based α-wave regulation has been used in the treatment of anxiety disorders, utilizing personalized reward mechanisms to optimize NF training. Studies show that increasing α-wave activity in the frontal lobe can reduce anxiety symptoms, and reinforcement learning-based NF is used to achieve this by providing individualized rewards for successful regulation (Ros et al., 2010; Enriquez-Geppert et al., 2017). DRL helps in dynamically adjusting the reward structure based on real-time EEG signals, which allows patients to achieve better outcomes through a tailored training process (Hammond, 2005). Personalized α-wave NF has been shown to significantly enhance relaxation and reduce anxiety compared to fixed-reward approaches, as the training targets individual - specific brain dynamics (Gevensleben et al., 2014).

3.3 Virtual Reality (VR)

Combining EEG with VR environments has been used to create immersive and personalized cognitive training experiences. By using EEG signals, personalized feedback can be delivered in real-time, thereby enhancing the effectiveness of the VR training (Bouchard et al., 2012). The immersive nature of VR, coupled with EEG-based personalized feedback, has been shown to improve user engagement and task performance. For example, DRL algorithms have been used to adjust VR scenarios in response to the user's cognitive state, measured through EEG signals, to provide an optimal level of challenge and reward. Such tailored interventions are particularly beneficial in treating anxiety disorders, where the immersive VR environment can simulate real-world situations while the EEG-based feedback helps the individual manage stress responses in real-time.

4 PERSONALISED REWARD

4.1 Deep Q-Networks and Personalized Rewards

DQN have been widely used for implementing personalized rewards in NF training, enabling individualized learning experiences based on each user's performance. In DQN, the agent learns to take actions that maximize cumulative rewards through approximating the optimal action-value function with deep neural networks (Mnih et al., 2015). This framework allows for the dynamic adjustment of rewards to better suit individual differences in cognitive training, thereby improving engagement and overall learning outcomes (Silver et al., 2018).

In the context of NF, DQN can be used to model complex reward structures that reflect the changing needs of participants during training. For example, personalized reward functions can be used to enhance the relevance and saliency of the NF signals provided, which has been shown to significantly improve motivation and training effectiveness (Enriquez-Geppert et al., 2017). By tailoring the reward system to the user's progress, DQN-based interventions can address the limitations of fixed reward strategies, ensuring that the feedback provided aligns closely with each individual's learning trajectory.

4.2 Application of Policy Gradient Methods in Personalized Rewards

Policy gradient methods, such as PPO and A3C, offer significant advantages for real-time adaptive interventions in NF training by optimizing the agent's policy directly through gradient ascent (Schulman et al., 2017). Unlike value-based methods like DQN, policy gradient methods allow for continuous action spaces and are particularly well-suited for environments where high adaptability is needed to accommodate individual differences (Mnih et al., 2016).

PPO and A3C have been used effectively in NF to adjust training tasks in real-time based on individual performance metrics. For example, PPO has been applied to optimize task difficulty and feedback parameters during NF sessions, ensuring that each participant receives a training experience tailored to their cognitive state (Watanabe et al., 2018). A3C, with its capability to use multiple agents concurrently, enables rapid learning and adaptation, making it ideal for adjusting personalized rewards in NF settings (Schulman et al., 2017). This capability ensures that users remain engaged and that the intervention remains effective over time, even as their performance fluctuates.

Moreover, policy gradient methods offer the flexibility to incorporate more complex reward structures, such as those involving physiological signals (e.g., heart rate variability), to provide a more comprehensive and individualized NF experience (Silver et al., 2018; Kothgassne et al., 2022). This flexibility allows for a holistic approach to cognitive training, where various bio-signals are considered in reward computation to enhance the efficacy of the training.

5 EXPERIMENTAL VALIDATION

5.1 Experimental Design and Results

Recent research has validated the effectiveness of personalized reward mechanisms in cognitive training, both in laboratory and clinical settings. For example, Enriquez-Geppert, Huster, and Herrmann (2019) conducted a randomized controlled trial (RCT) examining EEG-based NF for individuals with ADHD. In the experimental group, participants received personalized rewards based on their ability to regulate brain activity, specifically focusing on the TBR. The control group received fixed rewards. The personalized reward group demonstrated significant improvements in attention and executive functioning compared to the control group, highlighting the importance of real-time, individualized feedback for cognitive enhancement.

Watanabe, Sasaki, Shibata, and Kawato (2017) explored the use of personalized rewards in fMRI-based neurofeedback interventions for anxiety disorder patients. Using DRL algorithms, the reward structure was dynamically adapted based on the participants' ability to modulate brain activity in anxiety-related regions. The experimental group, which received personalized feedback, showed greater neural regulation and reduced anxiety symptoms compared to the control group, which received non-adaptive feedback.

A study by Bhargava, O'Shaughnessy, and Mann (2020) introduced a novel approach using RL in EEG-based NF. The authors designed a DQN system that modulated audio feedback in real-time based on brainwave activity, aiming to enhance participants' meditative states. Their results demonstrated that the personalized reward system led to significant improvements in participants' brain states compared to conventional NF systems, with faster convergence toward optimal outcomes. This further supports the utility of personalized feedback in improving the effectiveness of NF interventions.

Additionally, Tripathy et al. (2024) investigated the use of RL to optimize real-time interventions and personalized feedback using wearable sensors. The study demonstrated how the system used RL to dynamically adjust interventions based on real-time physiological data from wearable sensors, providing personalized feedback that was more responsive to the user's needs. This approach led to improved cognitive outcomes and greater user engagement in self-monitoring and health management tasks, further highlighting the benefits of personalized reward mechanisms.

These studies collectively demonstrate that personalized reward mechanisms provide significant benefits across various cognitive training applications, from ADHD treatment to meditation and real-time health management, by tailoring feedback to the individual's needs, leading to superior cognitive outcomes compared to fixed rewards.

5.2 Comparison with Fixed Reward Mechanisms

Personalized reward mechanisms have consistently proven to be more effective than fixed reward mechanisms in enhancing cognitive performance. Traditional RL, which uses fixed rewards, provides identical reinforcement regardless of individual performance, leading to reduced engagement and motivation over time (Sutton & Barto, 2018). By contrast, personalized rewards adapt dynamically to the learner's progress, offering feedback that is more meaningful and aligned with their specific abilities, which results in improved cognitive outcomes (Silver et al., 2018).

Tripathy et al. (2024) conducted a study comparing personalized and fixed reward mechanisms using wearable sensors in real-time intervention systems. The findings revealed that participants receiving personalized feedback exhibited significantly greater improvements in cognitive function and engagement levels compared to those receiving fixed rewards. The authors emphasized that personalized rewards, which adjust dynamically based on physiological and performance data, created a more engaging and effective learning environment.

In the context of DRL, Mnih et al. (2015) demonstrated that personalized rewards facilitated faster convergence to optimal policies. Their study used DQN to compare personalized and fixed rewards in simulated environments. The personalized reward group achieved higher performance levels in complex tasks as the feedback was more closely aligned with their learning trajectory. This adaptability allowed agents to learn more efficiently, reinforcing the advantages of personalized rewards for optimizing training outcomes.

In NF applications, personalized rewards have been found to promote greater neural plasticity and behavioral improvements compared to fixed rewards. Enriquez-Geppert, Huster, and Herrmann (2017) reported that participants who received personalized NF exhibited enhanced neuroplastic changes, such as increased connectivity between targeted brain regions. These neural changes were not observed in the group receiving fixed rewards, highlighting the superiority of personalized feedback in promoting adaptive changes in brain function.

6 CHALLENGES

6.1 Complexity of Data Processing and Model Design

The complexity of real-time EEG data processing and the computational demands of deep learning models pose significant challenges in implementing personalized reward mechanisms in cognitive training. Processing EEG signals in real-time requires precise temporal analysis and advanced algorithms to extract meaningful features, which can be computationally intensive. Sharma and Meena (2024) highlight emerging trends in EEG signal processing, particularly in noise reduction, artifact removal, and feature extraction, which are critical for enhancing data quality in real-time systems. These processes must handle various sources of noise and artifacts, such as eye movements and muscle contractions, which complicate the accurate detection of neural signals (Sharma & Meena, 2024). Advanced filtering techniques and robust preprocessing steps are essential to maintain signal integrity, but they also increase the computational load.

Additionally, DRL models used for personalized NF require substantial computational power due to their multi-layered architectures. The use of convolutional and recurrent neural networks in these models adds to the computational burden, making real-time adaptation challenging, particularly in resource-constrained environments (Mnih et al., 2015). Optimization techniques, such as model pruning or quantization, may help reduce latency, but achieving real-time performance remains a significant hurdle (Schulman et al., 2017).

6.2 Adaptability

Personalized reward mechanisms are designed to address individual differences in neural functioning, but their adaptability has limitations when applied to diverse populations. While personalized rewards can tailor cognitive training to an individual's neural activity, their effectiveness may vary across different demographic groups, such as varying ages, cultural backgrounds, and cognitive abilities. Enriquez-Geppert et al. (2019) demonstrated that personalized rewards enhance cognitive training outcomes, but noted that their adaptability is constrained by the variability in neural responses across individuals. This variability becomes particularly challenging in populations with distinct neurological conditions, such as ADHD or autism spectrum disorders, where standard personalization techniques may not be effective for everyone (Watanabe et al., 2018).

Furthermore, the process of calibrating personalized NF systems often requires extensive data collection and adaptation, limiting the scalability of these interventions. The trade-off between personalization and generalization remains an area of concern, particularly when attempting to develop systems that can cater to larger, more diverse populations (Silver et al., 2018). Future research should explore how these reward systems can be made more adaptive and inclusive while maintaining their personalized approach.

6.3 Future Research Directions

Future research should focus on integrating additional physiological signals, such as heart rate, skin conductance, and respiration, into personalized reward systems to provide a more comprehensive assessment of an individual's physiological state. Incorporating multimodal data sources alongside EEG could improve the robustness of the system and enable more accurate feedback mechanisms (Ros et al., 2013). For instance, combining EEG data with other bio-signals may allow for a more nuanced interpretation of an individual's cognitive and emotional states, thereby enhancing the effectiveness of personalized NF interventions.

Additionally, the rise of wearable technology provides opportunities for real-time monitoring in non-clinical settings, as highlighted by Tripathy et al. (2024). Wearables equipped with sensors that can capture various physiological parameters offer a way to extend personalized NF systems beyond clinical environments, potentially increasing accessibility and usability.

Advanced DRL techniques, such as meta-RL, could also play a crucial role in enhancing the adaptability of personalized reward systems. Meta-RL enables models to learn more quickly from fewer data points, which could reduce the calibration time required for personalized NF (Schulman et al., 2017). This approach may also facilitate the development of systems that are more responsive to individual differences, improving both the scalability and effectiveness of personalized interventions.

In addition to technical advancements, ethical considerations related to the use of personalized reward mechanisms must be addressed. Issues such as data privacy, the potential for unintended psychological effects, and the broader implications of highly personalized interventions should be carefully examined to ensure these systems are safe and ethically sound.

7 CONCLUSION

Personalized reward mechanisms play a pivotal role in enhancing the effectiveness of DRL-driven cognitive training. By dynamically adjusting rewards based on real-time user performance and physiological signals, personalized rewards provide more tailored and engaging feedback, significantly improving the efficacy of cognitive training interventions compared to traditional fixed reward systems (Silver et al., 2018; Tripathy et al., 2024). Personalized rewards ensure that tasks are optimally challenging for each participant, promoting sustained engagement, continuous learning, and overall cognitive improvement (Sutton & Barto, 2018).

In NF applications, personalized rewards have been shown to enhance attention, memory, and executive functions, while also helping to alleviate symptoms of ADHD and anxiety (Enriquez-Geppert et al., 2019; Watanabe et al., 2017). The integration of

DRL algorithms allows for real-time adaptation, adjusting training tasks based on a user's physiological and behavioral responses. This approach enhances the overall effectiveness of the intervention by providing individualized and contextually relevant feedback (Mnih et al., 2015; Bhargava et al., 2020).

However, there are challenges in implementing personalized reward mechanisms, such as the complexity of processing real-time EEG data and the computational demands of DRL models (Sharma & Meena, 2024). Advanced signal processing techniques are required to manage noise and variability in EEG signals, while DL models need optimization to reduce computational latency in real-time applications. Additionally, making these systems adaptable across diverse populations with varying neurological conditions remains an ongoing challenge, with current approaches often requiring extensive calibration to achieve effective personalization (Watanabe et al., 2017).

Future research should focus on integrating additional physiological signals, such as heart rate and skin conductance, into personalized NF systems to create more holistic feedback mechanisms (Ros et al., 2013). Advances in wearable technology could support real-time monitoring of multiple physiological parameters, broadening the scope of personalized cognitive training outside clinical settings (Tripathy et al., 2024). Moreover, exploring advanced DRL techniques, such as Meta-RL, could further enhance adaptability, enabling systems to learn from fewer data points and reduce calibration time (Schulman et al., 2017).

The potential of personalized reward mechanisms in cognitive training is immense. As the field progresses, addressing the challenges of model complexity, data processing, and adaptability will be crucial to fully realizing the benefits of personalized cognitive training. Ultimately, the integration of personalized rewards in DRL-driven interventions holds the promise of transforming cognitive enhancement and mental health treatments, making them more effective, individualized, and engaging for a wide range of user.

REFERENCES

Arns, M., Clark, C. R., Trullinger, M., DeBeus, R., Mack, M., & Aniftos, M. (2020). Neurofeedback and attention-deficit/hyperactivity-disorder (ADHD) in children: Rating the evidence and proposed guidelines. Applied Psychophysiology and Biofeedback, 45(2), 39-48. https://doi.org/10.1007/s10484-020-09455-2

Arns, M., Heinrich, H., & Strehl, U. (2014). Evaluation of neurofeedback in ADHD: The long and winding road. *Biological Psychology*, *95*, 108-115. https://doi.org/10.1016/j.biopsycho.2013.11.013

Bhargava, A., O'Shaughnessy, K., & Mann, S. (2020). A novel approach to EEG Neurofeedback via reinforcement learning. *2020 IEEE SENSORS*. https://doi.org/10.1109/sensors47125.2020.9278871

Bouchard, S., Bernier, F., Boivin, É., Morin, B., & Robillard, G. (2012). Using biofeedback while immersed in a stressful Videogame increases the effectiveness of stress management skills in soldiers. *PLoS ONE*, *7*(4), e36169. https://doi.org/10.1371/journal.pone.0036169

Chi, Y. M., Jung, T., & Cauwenberghs, G. (2010). Dry-contact and Noncontact Biopotential electrodes: Methodological review. *IEEE Reviews in Biomedical Engineering*, *3*, 106-119. https://doi.org/10.1109/rbme.2010.2084078

DeBettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A., & Turk-Browne, N. B. (2015). Closed-loop training of attention with real-time brain imaging. *Nature Neuroscience*, *18*(3), 470-475. https://doi.org/10.1038/nn.3940

D'Esposito, M. (2008). Faculty opinions recommendation of improving fluid intelligence with training on working memory. *Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature*. https://doi.org/10.3410/f.1109007.566174

Enriquez-Geppert, S., Huster, R. J., & Herrmann, C. S. (2017). EEG-neurofeedback as a tool to modulate cognition and behavior: A review tutorial. *Frontiers in Human Neuroscience*, *11*.
https://doi.org/10.3389/fnhum.2017.00051

Enriquez-Geppert, S., Smit, D., Pimenta, M. G., & Arns, M. (2019). Neurofeedback as a treatment intervention in ADHD: Current evidence and practice. *Current Psychiatry Reports*, *21*(6). https://doi.org/10.1007/s11920-019-1021-4

Garcia Pimenta M, Brown T, Arns M, Enriquez-Geppert S. Treatment Efficacy and Clinical Effectiveness of EEG Neurofeedback as a Personalized and Multimodal Treatment in ADHD: A Critical Review. Neuropsychiatr Dis Treat. 2021 Feb 25;17:637-648. doi: 10.2147/NDT.S251547.

Gevensleben, H., Kleemeyer, M., Rothenberger, L. G., Studer, P., Flaig-Röhr, A., Moll, G. H., Rothenberger, A., & Heinrich, H. (2013). Neurofeedback in ADHD: Further pieces of the puzzle. *Brain Topography*, *27*(1), 20-32. https://doi.org/10.1007/s10548-013-0285-y

Hammond, D. C. (2005). Neurofeedback treatment of depression and anxiety. *Journal of Adult Development*, *12*(2-3), 131-137. https://doi.org/10.1007/s10804-005-7029-5

Kothgassner, O. D., Goreis, A., Bauda, I., Ziegenaus, A., Glenk, L. M., & Felnhofer, A. (2022). Virtual reality biofeedback interventions for treating anxiety. *Wiener klinische Wochenschrift*, *134*(S1), 49-59. https://doi.org/10.1007/s00508-021-01991-z

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., ... & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. *International Conference on Machine Learning*, 1928-1937.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529-533. https://doi.org/10.1038/nature14236

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347. https://arxiv.org/abs/1707.06347

Ros, T., Munneke, M. A., Ruge, D., Gruzelier, J. H., & Rothwell, J. C. (2010). Endogenous control of waking brain rhythms induces neuroplasticity in humans. *European Journal of Neuroscience*, *31*(4), 770-778. https://doi.org/10.1111/j.1460-9568.2010.07100.x

Sharma, R., & Meena, H. K. (2024). Emerging trends in EEG signal processing: A systematic review. *SN Computer Science*, *5*(4). https://doi.org/10.1007/s42979-024-02773-w

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., Van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of go without human knowledge. *Nature*, *550*(7676), 354-359. https://doi.org/10.1038/nature24270

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., Weiskopf, N., Blefari, M. L., Rana, M., Oblak, E., Birbaumer, N., & Sulzer, J. (2016). Closed-loop brain training: The science of neurofeedback. *Nature Reviews Neuroscience*, *18*(2), 86-100. https://doi.org/10.1038/nrn.2016.164

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

Tripathy, J., Balasubramani, M., Rajan, V. A., S, V., Aeron, A., & Arora, M. (2024). Reinforcement learning for optimizing real-time interventions and personalized feedback using wearable sensors. *Measurement: Sensors*, *33*, 101151. https://doi.org/10.1016/j.measen.2024.101151

Watanabe, T., Sasaki, Y., Shibata, K., & Kawato, M. (2018). Advances in fMRI real-time Neurofeedback. *Trends in Cognitive Sciences*, *22*(8), 738. https://doi.org/10.1016/j.tics.2018.05.007

Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3-4), 279-292.