Summary of the Feedback Session for Our Final Project in the "Deep Reinforcement Learning" Course (Spring Semester 2023) on August 31st, 2023.

On the afternoon of Thursday, August 31st, our dedicated group—comprising Felix Hammer, Jonas Bieber, and Benjamin Fricke—gathered outside the imposing edifice at Wachsbleiche 27, Osnabrück. It was a few minutes shy of 13:30, and the diffident sun timidly pierced through the rain-heavy clouds. A serene hush enveloped the stairwell, and our footsteps reverberated through the desolate hallways. Our destination was the office of Mr. Leon Schmid, nestled in room 402a of the edifice officially named the "Institut für Kognitionswissenschaften," affectionately known among the initiated as "Wachsbleiche." Our mission? To subject our initial forays into the world of deep reinforcement learning, our final project, to the rigorous scrutiny of critical analysis. Our excitement was palpable.

Upon entering Mr. Schmid's office, we were greeted by a solitary figure, and he gestured toward a well-worn couch. With bated breath, we unfurled our laptops, each click of the keyboard punctuating the tension in the room. Almost breathlessly, we embarked on our presentation, a narrative we had rehearsed in the quietude of evening contemplations. With unwavering zeal, we recounted the project's enchanting inception, our tentative explorations into the realm of the unknown, the exhilarating triumphs, the crushing setbacks, and our resilient comebacks. For approximately 25 minutes, we relived our odyssey, punctuated only by sporadic inquiries from our esteemed interlocutor. Exhausted but hopeful, we awaited Mr. Schmid's verdict. To our immense delight, his judgment was unequivocally positive. A smile graced his lips as he acknowledged the charming innocence of attempting to solve Sudoku puzzles with deep reinforcement learning, and it was evident that he shared our passionate commitment. However, he astutely pointed out the daunting challenge posed by the expansive action space. We reassured him that we were already addressing this issue through action masking. In that moment, it felt like all had been said.

Following the verdict, Mr. Schmid continued to give us a few valuable pieces of advice, which we would like to list below, so that we do not forget them:

1. We had grappled with the formulation of a Sudoku Markov Decision Process (MDP) but lacked a clear solution at the time. In the clarity that emerges when one endeavors to elucidate their tangled thoughts to a seasoned mentor, it became apparent that our approach—terminating a game upon the agent's first mistake—was suboptimal. Mr. Schmid suggested allowing the agent to make multiple errors.
2. Mr. Schmid recommended exploring transformer architectures alongside the neural architectures we had already ventured into. While this idea had crossed our minds, his encouragement rekindled our interest.
3. He provided valuable guidance on the potential use of visualizations like integrated gradients or linear probes.

Just as a hearty meal satisfies the body, our spirits were nourished and invigorated by this rich input. Our parting was marked by courteous words, and as we left, not a single utterance passed between us. Lost in contemplative reverie, we traversed the sunlit path. An unexpected surge of motivation washed over us, and we joyfully embraced, already planning to reconvene later in the afternoon to discuss and implement the wealth of insights we had gained. With hearts full of happiness, we embarked on our separate paths. The sun continued to shine, a beacon of optimism for our endeavors ahead.