Research Assignment Evaluation Questions

Why did you pick the particular project?

Project 1 (RL+LLM Reward Structures): I chose this because it seemed like the most straightforward way to generate meaningful research insights within the time constraints. The question of hard vs soft rewards is fundamental to RL, and I could implement controlled experiments to get clear comparative data.

Project 2 (Neural Flappy Bird): Honestly, I picked this because I really love Flappy Bird as a game, and the idea of teaching a neural network to understand its physics through pure visual observation seemed incredibly cool. The challenge of extending Neural Atari to a different domain was appealing, even though I underestimated the technical difficulties.

If you had more compute/time, what would you have done?

RL+LLM Project:

- Train for 1000+ steps instead of just 100 to reach true convergence
- Implement proper train/test splits for rigorous evaluation
- Replace Gemini API with deterministic local evaluation models
- Test creative domains like poetry and jokes where perplexity rewards should actually work
- Run multiple seeds for statistical significance testing

Flappy Bird Project:

- Implement 3D VQ-VAE architectural modifications for better temporal consistency
- Collect training data focused exclusively on physics transitions (no steady-state flight)
- Add temporal consistency losses to penalize static generation
- Try alternative architectures like Transformer-based world models or Google's Genie 3 approach
- Train for 50,000+ steps to see if the static generation problem resolves with longer training

What did you learn in the project?

Technical Lessons:

- Reward function design must perfectly align with the optimization objective
- External API dependencies break training loops due to inconsistency and rate limiting
- Quantitative training metrics can mislead about actual model learning capabilities
- Manual data inspection reveals critical biases that automated analysis misses completely

Research Methodology Lessons:

- Negative results provide valuable scientific insights about approach limitations
- Research requires embracing failure as data rather than seeing it as wasted effort

- Sometimes architectural limitations require completely starting over rather than parameter tuning
- Data quality matters infinitely more than data quantity

Personal Lessons:

- My stubbornness kept me working on the Flappy Bird project for 20-25 hours despite early warning signs
- Research thinking differs fundamentally from engineering thinking
- Compatibility issues and infrastructure problems often determine success more than algorithmic sophistication

What surprised you the most?

The VQ-VAE static generation problem shocked me more than any of the RL failures. Despite feeding it training data with 99.9% dynamic frame pairs averaging 27 pixels of movement per frame, the model consistently generated static sequences with virtually no bird movement. I genuinely believed the visual physics learning approach would work and found the challenge incredibly engaging.

The Gemini API breakdown during RL training was also unexpected. I didn't anticipate that external evaluation would introduce such fatal inconsistencies that it would completely break the learning process.

If you had to write a paper on the project, what else needs to be done?

For Publication-Quality Research:

1. **Statistical Rigor**: Multiple random seeds, proper train/test splits, statistical significance testing across all experiments
2. **Baseline Comparisons**: Compare against established RL+LLM methods and standard VQ-VAE implementations
3. **Architectural Analysis**: Systematic investigation of why VQ-VAE fails at temporal physics - is it fundamental to the approach or fixable with modifications?
4. **Human Evaluation**: Replace broken API evaluation with proper human preference studies for dialogue tasks
5. **Theoretical Framework**: Develop formal criteria for when different reward structures should be applied to different task domains
6. **Extended Experiments**: Test the reward structure findings on more diverse tasks beyond just math and dialogue
7. **Reproducibility Package**: Clean, well-documented code that others can actually run and replicate
8. **Related Work**: Comprehensive literature review comparing our findings to existing RL+LLM and world model research

The current work provides valuable negative results and methodology insights, but needs significantly more experimental depth and theoretical grounding for academic publication standards.