TASK 2: Effect of Learning Rate on Reinforcement Learning Agent Performance in Tic-Tac-Toe

Multiple Independent Runs:

- 5 separate experiments were conducted for each learning rate
- Each experiment trained an agent for 10,000 games
- Why 5 runs? Single experiments can be misleading due to randomness
- Results were averaged across all 5 runs to get reliable conclusions

Learning Rates Tested:

We tested 5 different learning speeds (α values):

- α = 0.01: Very slow learning
- α = 0.2: Medium-slow learning
- α = 0.5: Medium learning
- α = 0.7: Medium-fast learning
- α = 0.99: Very fast learning

Training Opponents:

- 30% Optimal Teacher: Makes smart moves 30% of time, random moves 70% of time
- 90% Optimal Teacher: Makes smart moves 90% of time, random moves 10% of time

Graph Type 1: Final Performance Summary

What It Shows:

- X-axis: Different learning rates (α = 0.01, 0.2, 0.5, 0.7, 0.99)
- Y-axis: Final win rate after full training
- Single line with dots: Final performance of each learning rate
- Numbers above dots: Exact win percentage values

What It Tells Us:

- Direct comparison of which learning rate works best
- Final ranking from best to worst performer
- Clear winner identification at a glance
- Quantitative results with precise percentages

Graph Type 2: Learning Curves Over Time

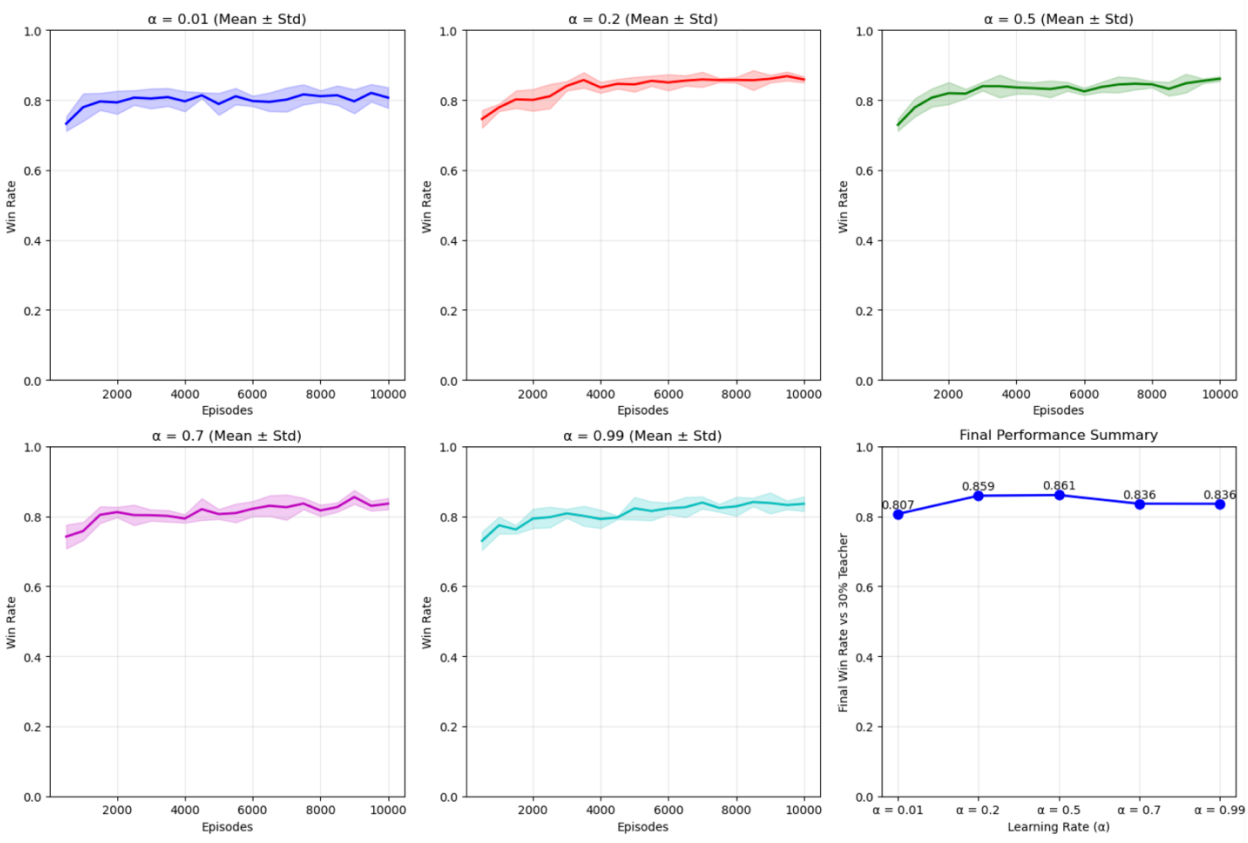
What It Shows:

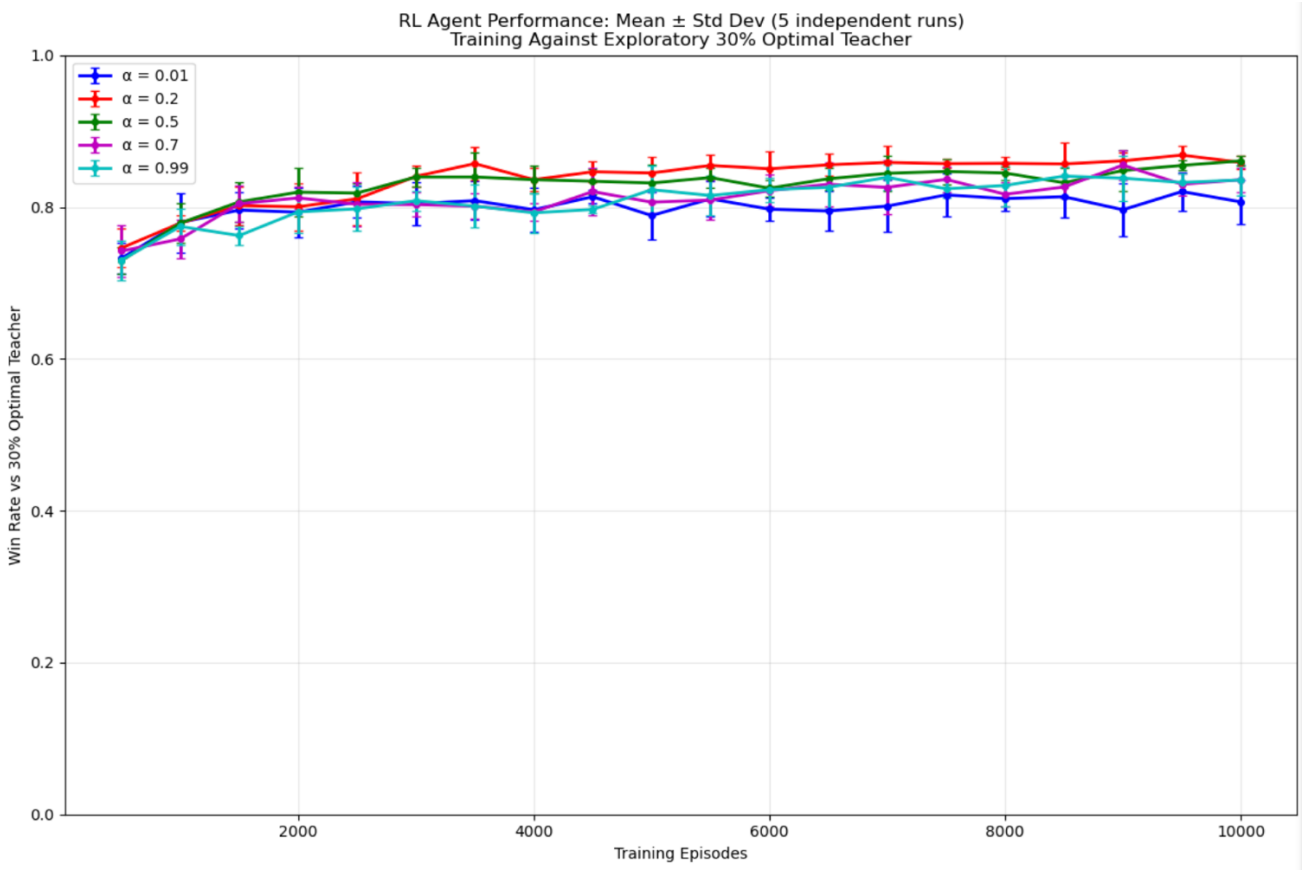
- X-axis: Training episodes (0 to 10,000 games)
- Y-axis: Win rate percentage (0% to 100%)
- Multiple colored lines: Each represents a different learning rate (α)
- Vertical error bars: Show variation across 5 independent runs

What It Tells Us:

- How learning progresses over time for each method
- When agents reach peak performance (usually around 2000-3000 games)
- Which learning rates learn faster vs slower
- Consistency of results (smaller error bars = more reliable)

1) 30% Optimal Player (Teacher)





Performance Results:

- α = 0.5: 86.1% win rate (highest)
- α = 0.2: 85.9% win rate (very close second)
- α = 0.7: 83.6% win rate
- α = 0.99: 83.6% win rate
- α = 0.01: 80.7% win rate (lowest)

Learning Dynamics:

- Fast convergence: All agents reach near-final performance by 2000-3000 episodes
- Stable plateau: Performance remains consistent after initial learning phase
- Minimal variance: Small error bars indicate predictable, repeatable results
- High baseline: Even worst performer (α = 0.01) achieves 80%+ win rate

Learning Rate Effects:

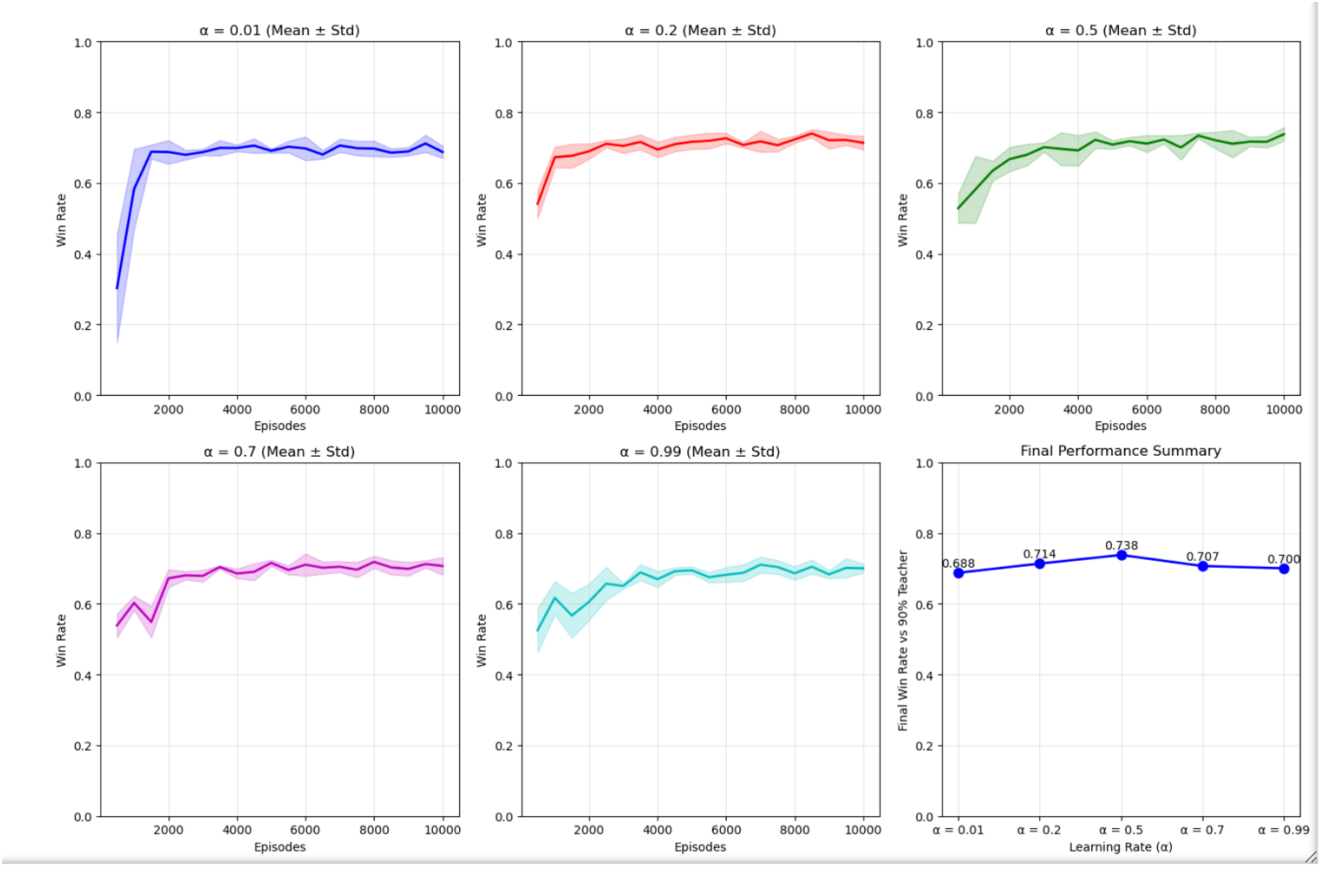
- α = 0.01: Shows characteristic slow learning but eventually reaches good performance
- α = 0.2: Rapid learning with excellent final performance
- α = 0.5: Quick convergence to highest performance level

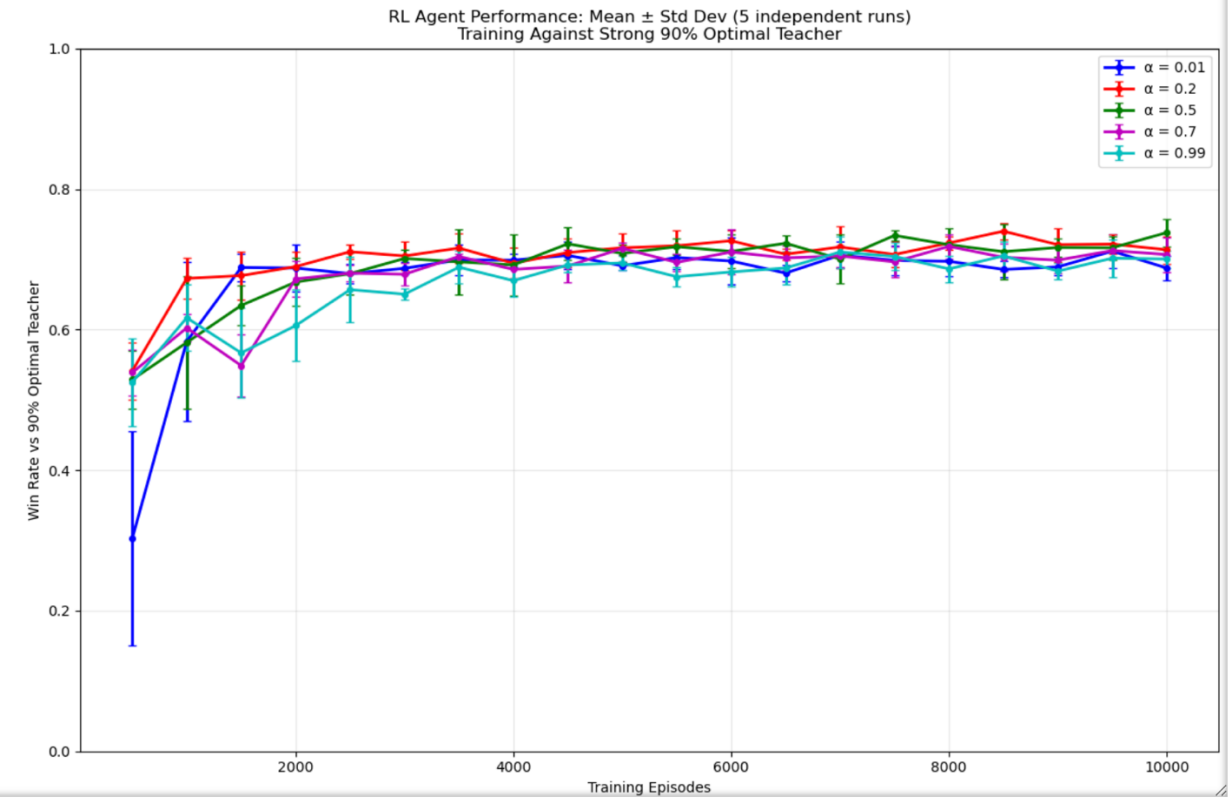
- α = 0.7: Good performance but slightly below optimal medium rates
- α = 0.99: Moderate performance, shows some instability

Key Insights:

- Forgiving environment: 70% random opponent moves make task relatively easy
- Ceiling effect: All learning rates can succeed against weak opposition
- Limited differentiation: Small performance gaps (5.4% range) between best/worst

2) 90% Optimal Player (Teacher)





Performance Results:

- α = 0.5: 73.8% win rate (clear winner)
- α = 0.2: 71.4% win rate
- α = 0.7: 70.7% win rate
- α = 0.99: 70.0% win rate
- α = 0.01: 68.8% win rate (significant gap)

Learning Dynamics:

- Challenging initial phase: All agents struggle early (30-50% win rates initially)
- Dramatic α = 0.01 curve: Starts at ~30%, climbs slowly to 68.8%
- Extended learning period: Requires full 10,000 episodes for convergence
- Higher variance: Larger error bars showing increased uncertainty

Learning Rate Effects:

- α = 0.01: Extremely slow start, massive improvement curve, but still lowest final performance
- α = 0.2: Steady improvement, good final performance
- α = 0.5: Best balance of learning speed and final performance
- α = 0.7: Moderate performance with some volatility
- α = 0.99: Inconsistent learning, prone to instability against strong opponent

Key Insights:

- Demanding environment: 90% optimal moves create sustained pressure
- True differentiation: Learning rate choice becomes critically important
- Tactical development: Agents must learn sophisticated strategies to compete

3. Comparative Analysis

Performance Impact of Opponent Strength:

Learning Rate (α)	30% Teacher Win Rate	90% Teacher Win Rate	Performance Drop
0.01	80.7%	68.8%	-11.9%
0.2	85.9%	71.4%	-14.5%
0.5	86.1%	73.8%	-12.3%
0.7	83.6%	70.7%	-12.9%
0.99	83.6%	70.0%	-13.6%

Learning Rate Sensitivity:

- 30% Teacher: Performance range = 5.4% (86.1% - 80.7%)
- 90% Teacher: Performance range = 5.0% (73.8% - 68.8%)

Optimal Learning Rate Identification:

- 30% Teacher: $\alpha = 0.5$ and $\alpha = 0.2$ perform virtually identically
- 90% Teacher: $\alpha = 0.5$ emerges as clear winner with 2.4% advantage over $\alpha = 0.2$

Learning Curve Characteristics:

30% Teacher:

- Smooth, predictable learning curves
- Quick convergence (2000-3000 episodes)
- Stable performance plateaus
- Low variance across runs

90% Teacher:

- More volatile learning trajectories
- Extended learning periods (full 10,000 episodes needed)
- Higher variance in results
- Clear separation between learning rate effectiveness

5. Practical Implications

Training Environment Design:

- 30% teacher: Good for initial learning and confidence building
- 90% teacher: Essential for developing competitive agents

Learning Rate Selection:

- α = 0.5 emerges as robust choice across both environments
- α = 0.2 performs well but shows more environment sensitivity
- **Very high (0.99) or low (0.01) rates show clear disadvantages

Agent Quality Assessment:

- 30% trained agents: Agents trained with 30% teacher might look like they're doing well, but they don't learn smart strategies.
- 90% trained agents: More likely to perform well against human players

Statistical Methodology:

- Error bars essential: Single runs can be misleading
- Multiple independent runs: Critical for reliable conclusions

6. Conclusions

1. Medium learning rates (α = 0.2-0.5) consistently outperform extreme values
2. Opponent strength dramatically affects both absolute performance and relative learning rate importance
3. Strong opponents are necessary for developing tactically competent agents
4. Statistical averaging across multiple runs is essential for reliable conclusions
5. α = 0.5 emerges as the most robust learning rate across different training conditions

This comprehensive analysis demonstrates how training environment difficulty fundamentally shapes both agent performance and the criticality of hyperparameter optimization in reinforcement learning systems.