

ML Model Performance Analysis

Overview

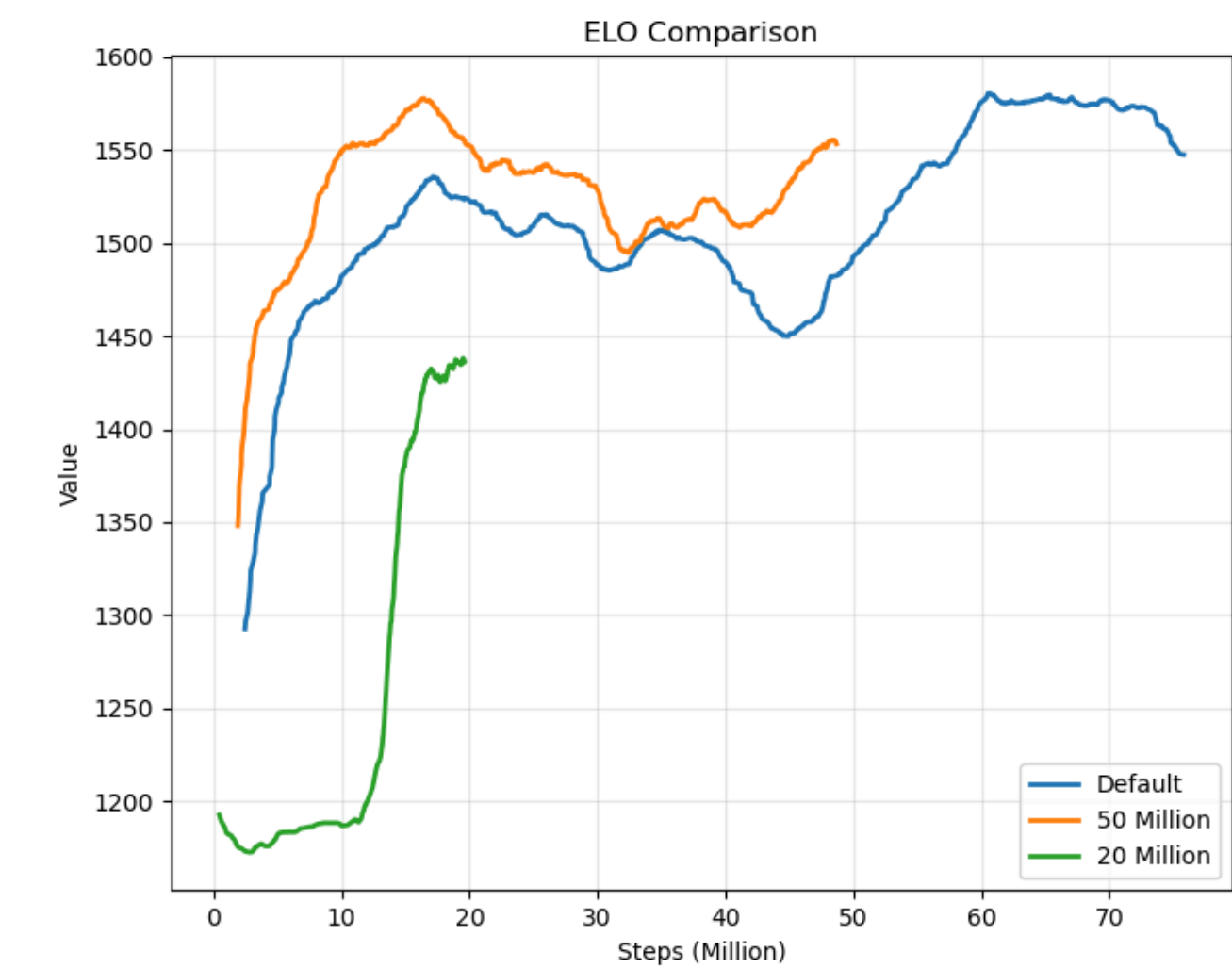
This analysis compares different Memory Sensor training steps against the base case, examining different level of steps.

Configurations

- Base Case (Default)
- Memory Sensor (50 million steps)
- Memory Sensor (20 million steps)

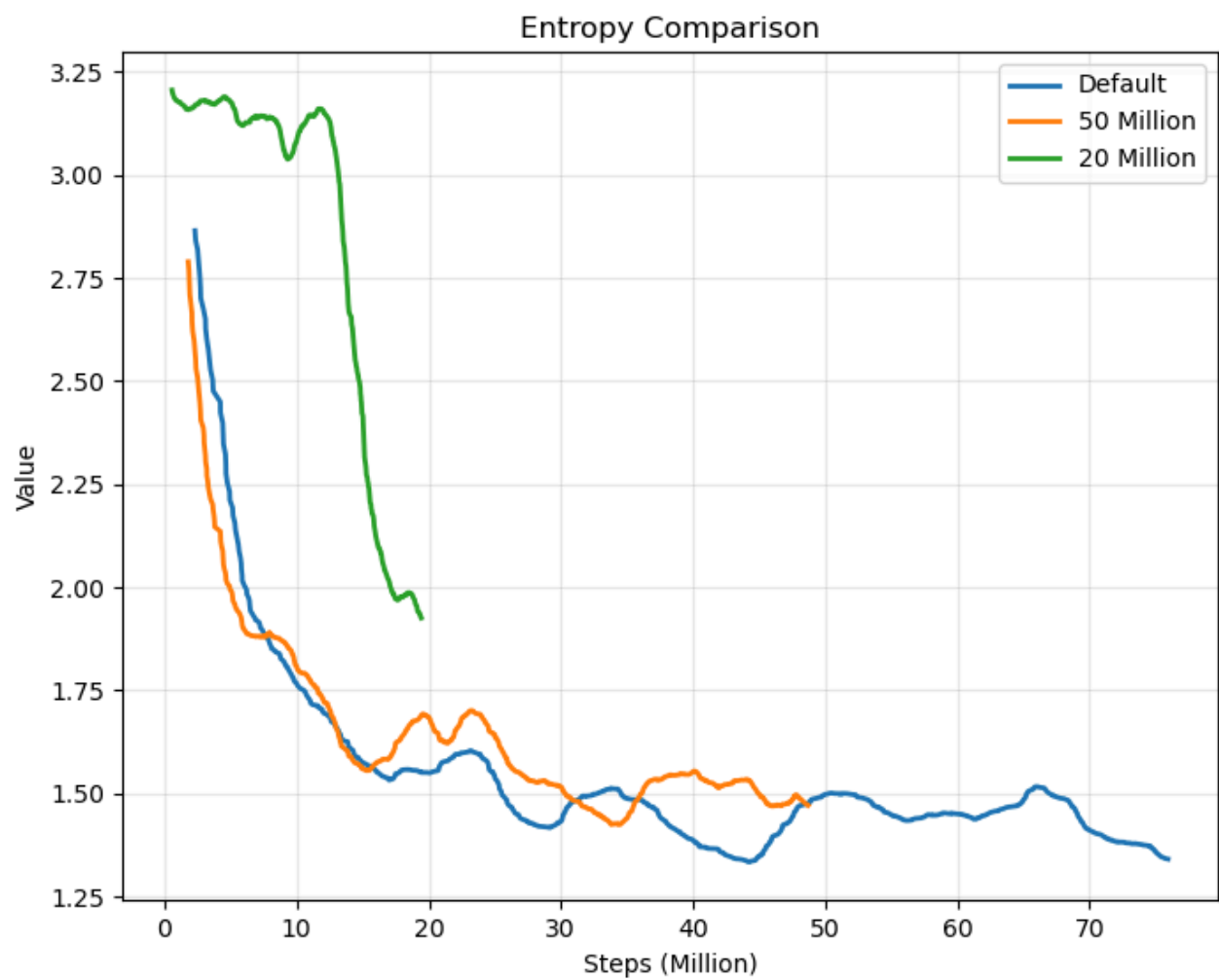
Metrics Analysis

ELO Rating



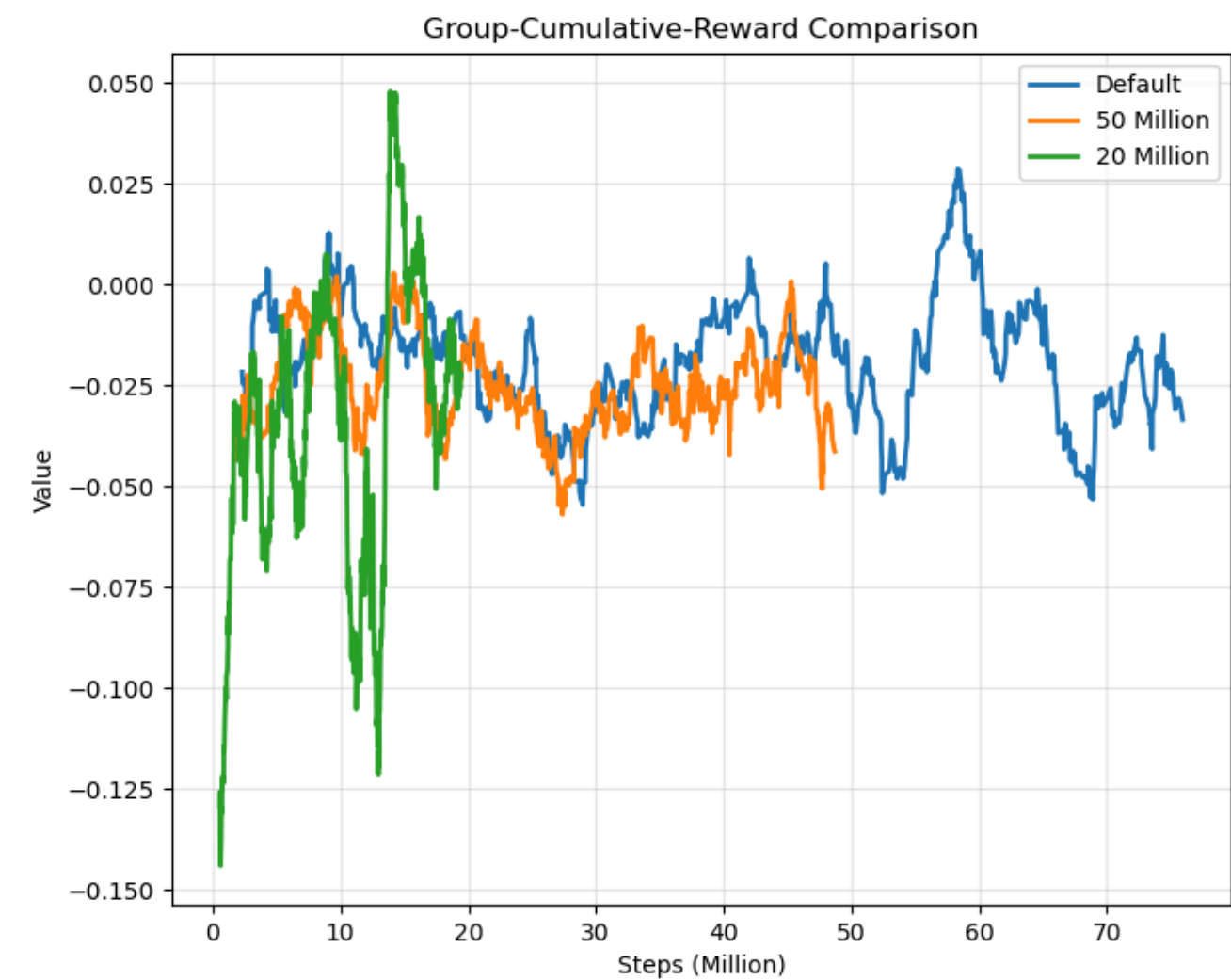
Configuration	Mean	Std Dev	Min	Max
Base Case	1506.60	66.53	1199.95	1608.24
Memory Sensor 50	1522.45	53.22	1194.28	1601.35
Memory Sensor 20	1280.55	109.63	1167.39	1459.09

Entropy



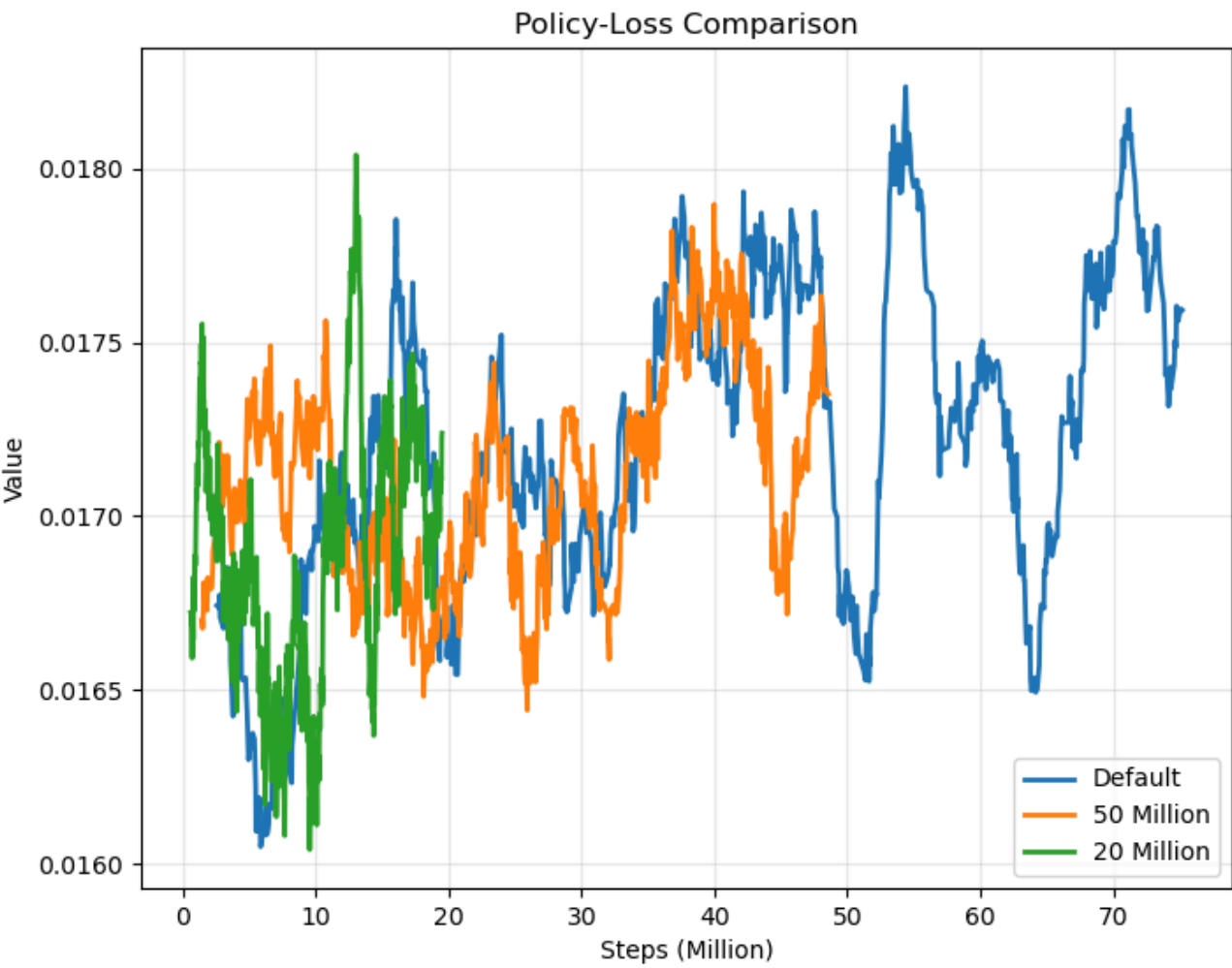
Configuration	Mean	Std Dev	Min	Max
Base Case	1.57	0.34	1.26	3.29
Memory Sensor 50	1.68	0.31	1.37	3.28
Memory Sensor 20	2.81	0.49	1.85	3.28

Group Cumulative Reward



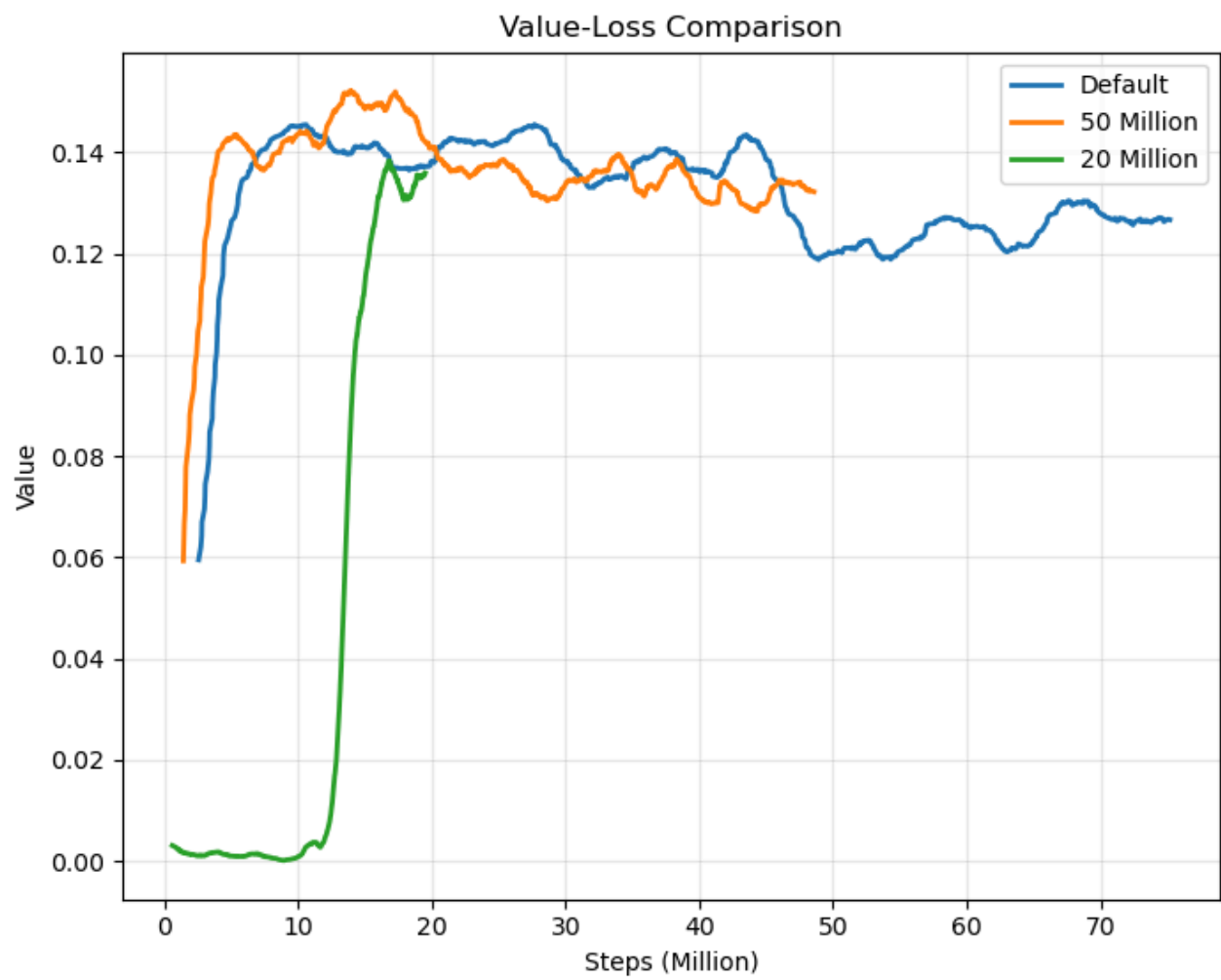
Configuration	Mean	Std Dev	Min	Max
Base Case	-0.019	0.098	-0.359	0.293
Memory Sensor 50	-0.026	0.095	-0.590	0.307
Memory Sensor 20	-0.036	0.137	-0.615	0.380

Policy Loss



Configuration	Mean	Std Dev	Min	Max
Base Case	0.017	0.002	0.010	0.027
Memory Sensor 50	0.017	0.002	0.011	0.028
Memory Sensor 20	0.017	0.002	0.011	0.026

Value Loss



Configuration	Mean	Std Dev	Min	Max
Base Case	0.130	0.026	0.000	0.160
Memory Sensor 50	0.134	0.022	0.000	0.165
Memory Sensor 20	0.044	0.058	0.000	0.150

Key Observations

1. Goals Scored

Fixed Delta Time: 0.01

Sensor Type	Goals Scored	Sensor Type	Goals Scored	Total Goals	Ratio
Base Case	111967	Memory Based 50	105574	217541	1.06055
Base Case	6518	Memory Based 20	5994	12512	1.08742

- **Description:**

This metric tracks the number of goals scored by agents using different sensor configurations during training. Comparing these values helps assess how memory-based sensors influence performance relative to the Base Case.

- **Evaluation:**

- Memory-Based 50 configuration scored 105,574 goals, slightly fewer than the Base Case (111,967). Despite this, the total goals scored by the combined configurations yielded a ratio of 1.06055, indicating an improvement in collaborative scoring.
- Memory-Based 20 configuration scored 5,994 goals, again slightly fewer than the Base Case (6,518), with a higher ratio of 1.08742, suggesting memory played a more significant role in this shorter training cycle.
- The results indicate that memory sensors influence gameplay dynamics differently depending on training duration. Further analysis of gameplay behavior and strategy adjustments is needed to understand these differences fully.

2. ELO

- **Description:**

ELO measures the relative performance of agents in competitive scenarios. A higher ELO indicates better adaptability, strategic decision-making, and overall success in the simulated environment.

- **Evaluation:**

- The Memory Sensor (50 million steps) configuration achieved a mean ELO of 1522.45, outperforming the Base Case (1506.60) and Memory Sensor (20 million steps) (1280.55).
- The lower standard deviation for Memory Sensor (50 million steps) (53.22) compared to Memory Sensor (20 million steps) (109.63) indicates greater consistency.
- The Base Case's relatively high minimum ELO suggests that default agents had a strong baseline but were outperformed in peak performance by the Memory Sensor configurations.

3. Entropy

- **Description:**
Entropy reflects the diversity of an agent's actions. Higher entropy indicates broader exploration, which is valuable during early training but may hinder policy convergence in later stages.
- **Evaluation:**
 - Memory Sensor (50 million steps) had slightly higher entropy (1.68) compared to the Base Case (1.57), suggesting enhanced exploration while still maintaining control.
 - Memory Sensor (20 million steps) showed significantly higher entropy (2.81), indicating excessive exploration that might have contributed to less stable training outcomes, as seen in its lower ELO and higher variance.

4. Group Cumulative Rewards

- **Description:**
Cumulative rewards measure the total rewards accumulated by agents. A close-to-zero value suggests balance in competitive scenarios, while larger deviations may indicate challenges or suboptimal strategies.
- **Evaluation:**
 - Base Case achieved a mean cumulative reward of -0.019, closely matched by Memory Sensor (50 million steps) with -0.026.
 - Memory Sensor (20 million steps) recorded -0.036, reflecting greater challenges in reward accumulation, likely due to its high entropy and less refined strategies.
 - The slightly higher variability in rewards for Memory Sensor configurations indicates that adding memory impacts the reward dynamics, particularly in earlier training phases.

5. Policy Loss

- **Description:**
Policy loss quantifies the adjustments made to the agent's policy during training. Lower values suggest more stable policy updates, which is desirable for convergence.
- **Evaluation:**
 - All configurations exhibited comparable mean policy loss (0.017), indicating stability in training across setups.
 - Memory Sensor configurations did not introduce significant instability, suggesting that the memory mechanism was effectively integrated into the training process.

6. Value Loss

- **Description:**

Value loss measures the accuracy of predicting future rewards. Lower loss suggests better predictions and a stronger understanding of the environment.

- **Evaluation:**

- Base Case had a mean value loss of 0.130, while Memory Sensor (50 million steps) was slightly higher at 0.134, showing comparable performance.
- Memory Sensor (20 million steps) recorded a significantly lower value loss (0.044), which might reflect simpler policies and less complexity in its learned value functions. However, this simplicity may have contributed to its poorer overall performance (ELO and rewards).