***Classification of Events***

**True Positives (TP) - Correct Predictions by LLM**

These are events that the LLM correctly predicted and also happened during the real Battle of Midway.

| **Event** | **Definition** |
| --- | --- |
| The battle was primarily a carrier-based naval engagement | Midway was a carrier-dominated battle, with aircraft playing a decisive role. |
| The defending side (USA) had an intelligence advantage | The U.S. broke Japanese codes, allowing them to anticipate and prepare for the attack. |
| The attacking force (Japan) aimed to seize a strategic island | Japan’s goal was to capture Midway Atoll, forcing the U.S. into a weaker position in the Pacific. |
| The defending side (USA) set up an ambush | The U.S. positioned carriers strategically to counter the Japanese assault. |
| Aircraft played the primary offensive role | The battle was won through carrier-launched dive bombers and torpedo planes. |
| The attacking force (Japan) had numerical superiority at the start | The Japanese Navy initially had more aircraft carriers and battleships. |
| The defending side (USA) suffered early losses but recovered | The U.S. lost aircraft and Yorktown was severely damaged before launching a decisive counterattack. |
| Multiple Japanese aircraft carriers were destroyed | The U.S. sank four Japanese carriers (*Akagi, Kaga, Soryu, and Hiryu*). |
| The battle was a turning point in the Pacific War | Midway marked the shift of momentum, placing Japan on the defensive. |
| The Japanese fleet retreated after suffering heavy losses | With four carriers destroyed, Japan abandoned its Midway invasion plan and withdrew. |

Total True Positives (TP): **10**

**False Positives (FP) - Incorrect Predictions by LLM**

These are events that the LLM predicted but did NOT actually happen in the real Battle of Midway.

| **Event** | **Why It Did NOT Happen?** |
| --- | --- |
| The U.S. launched a major counter-invasion after victory | The U.S. did not immediately invade Japanese-held islands after Midway. |
| Japan successfully landed troops on Midway Atoll | No Japanese troops landed, as their carriers were destroyed before they could launch an invasion. |
| The battle involved a significant submarine attack | While submarines played a role, Midway was mainly an aerial battle. |
| A ceasefire or temporary truce was declared mid-battle | The battle was fought until Japan’s retreat, with no negotiations. |
| Japan relied on radar technology for fleet defense | Japan had limited radar capabilities, relying mostly on scout planes. |
| The U.S. suffered equal carrier losses as Japan | The U.S. only lost one carrier (USS Yorktown), while Japan lost four. |
| The battle included large-scale kamikaze attacks | Kamikaze tactics were not used in 1942, as Japan only started them in 1944. |
| The battle involved direct battleship engagements | The battle was entirely fought with aircraft, with no battleship duels. |
| Japan had major reinforcements that never arrived | Japan committed all available carriers to the battle, with no reinforcements available. |

Total False Positives (FP): **9**

**False Negatives (FN) - Missed Predictions by LLM**

These are real events that happened during the Battle of Midway but were NOT predicted by the LLM.

| **Event** | **Definition** |
| --- | --- |
| The U.S. deliberately set a trap using false radio transmissions | The U.S. broadcasted misleading messages about Midway’s water supply, tricking Japan into confirming its target. |
| Japanese scout planes failed to detect U.S. carriers in time | A crucial Japanese scout plane was delayed, preventing Japan from spotting U.S. forces early enough. |
| U.S. torpedo bombers suffered heavy losses before dive bombers struck | Early torpedo bomber squadrons were wiped out, but their distraction set up the decisive dive bomber attack. |
| Weather conditions impacted visibility and engagements | The battle was influenced by cloud cover, affecting aircraft operations. |
| Admiral Yamamoto commanded Japan’s fleet from afar | Yamamoto planned the battle but was too far away to respond effectively. |
| The Japanese fleet was divided, making coordination difficult | Japan’s main force and supporting ships were spread out, delaying reinforcements. |
| U.S. aircraft had tactical advantages in survivability | While some Japanese planes were superior, U.S. aircraft had better coordination and defensive formations. |
| Midway was Japan’s last major offensive in the Pacific | After Midway, Japan never regained the strategic initiative and shifted to defensive warfare. |
| Japanese survivors retreated under cover of darkness | After their carriers were lost, Japanese forces withdrew at night to avoid further losses. |
| U.S. dive bombers struck at the exact moment Japanese planes were rearming | The most decisive moment occurred when U.S. bombers hit Japanese carriers while their aircraft were refueling and rearming. |

Total False Negatives (FN): **10**

***Final Results***

| **Metric** | **Value** |
| --- | --- |
| Total True Positives (TP) | **10** |
| Total False Positives (FP) | **9** |
| Total False Negatives (FN) | **10** |
| Precision | **52.63%** |
| Recall | **50.00%** |
| F1-Score | **51.24%** |

Critical Analysis of LLM’s Prediction

1. Major Issue: The LLM Predicted the Wrong Winner

* The LLM incorrectly assumed Japan won, when in reality, the USA decisively won the battle.
* Japan lost four carriers, effectively ending its ability to conduct major offensive operations.

2. Strengths in LLM’s Predictions

Correctly identified key aspects of the battle:

* Carrier warfare was central to the battle.
* U.S. intelligence (codebreaking) was a major advantage.
* Japan started with numerical superiority but suffered a crushing defeat.
* The battle was a turning point in the Pacific War.

3. Weaknesses in LLM’s Predictions

Overestimated Japan’s ability to sustain the battle:

* Japan’s logistical and intelligence failures were not fully accounted for.
* The U.S. suffered early losses, but not enough to shift the outcome.

Missed key tactical moments:

* Japan’s scouting failures played a critical role in their defeat.
* U.S. dive bombers struck at a critical moment when Japan was rearming.

Invented historical inaccuracies:

* No Japanese troops landed on Midway.
* No ceasefire or diplomatic resolution was attempted.
* No major battleship engagements occurred.

***Final Verdict***

* The LLM’s prediction was weak (~51.24% F1-score) due to incorrect outcome prediction and historical inaccuracies.
* Improving recall (identifying more real-world details) and correcting historical errors would significantly improve future assessments.