***Classification of Events***

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

These events were correctly predicted by the LLM and also happened in the real Cuban Missile Crisis.

| **Event** | **Definition** |
| --- | --- |
| Presence of nuclear-capable missiles as a central issue | The crisis revolved around the Soviet deployment of nuclear missiles in Cuba, which the U.S. perceived as a direct threat. |
| Naval blockade as a response strategy | The U.S. imposed a naval blockade ("quarantine") around Cuba to prevent further Soviet military shipments. |
| High-stakes nuclear confrontation | The crisis marked the closest moment to nuclear war between the U.S. and the Soviet Union during the Cold War. |
| Diplomatic negotiations played a decisive role | A deal was reached between Kennedy and Khrushchev, avoiding military escalation. |
| Use of intelligence and aerial surveillance | U-2 spy plane missions confirmed the presence of Soviet missile sites in Cuba, prompting the crisis. |
| Backchannel diplomacy between U.S. and Soviet Union | Secret negotiations between Robert Kennedy and Soviet Ambassador Anatoly Dobrynin helped resolve the crisis. |
| Strategic alliances influenced decision-making | The U.S. had NATO support, while the Soviets had Cuba as an ally, influencing strategic calculations. |
| Mutual agreement to de-escalate | The crisis was resolved when the U.S. agreed to remove missiles from Turkey in exchange for the Soviet withdrawal from Cuba. |
| Soviet ships initially approached blockade but later turned back | The world watched as Soviet cargo ships carrying military equipment neared the blockade before ultimately stopping. |
| The crisis led to long-term arms control agreements | The crisis directly influenced the signing of the Nuclear Test Ban Treaty (1963) and later arms reduction talks. |

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 Cuban Missile Crisis.

| **Event** | **Why It Did NOT Happen?** |
| --- | --- |
| Direct military confrontation between U.S. and Soviet forces | While tensions were extremely high, there was no direct military engagement between U.S. and Soviet forces. |
| Large-scale invasion of Cuba by U.S. | The U.S. considered an invasion (e.g., *Operation Ortsac*), but it never happened. |
| Nuclear weapons were launched or detonated | The crisis remained a standoff, with no nuclear weapons used. |
| Soviet submarines launched an attack | A Soviet submarine nearly launched a nuclear torpedo but was stopped by a second officer’s intervention. However, no actual attack occurred. |
| Cuba played an independent diplomatic role | While Cuba was a key location, Fidel Castro was not directly involved in the resolution of the crisis—it was mainly a U.S.-Soviet negotiation. |
| The U.S. lost control of the situation | Despite the crisis's intensity, the U.S. maintained strategic control and avoided direct war. |
| Soviet military forces engaged in combat with U.S. forces | The Soviets never engaged in combat with U.S. forces during the crisis. |
| Missile strikes occurred in Europe | There were fears of escalation, but no Soviet missile attacks on Turkey, Italy, or other NATO territories took place. |

Total False Positives (FP): **8**

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

These are real events that happened in the Cuban Missile Crisis but were NOT predicted by the LLM.

| **Event** | **Definition** |
| --- | --- |
| The U.S. had already placed nuclear missiles in Turkey and Italy before the crisis | The Soviet decision to place missiles in Cuba was a response to U.S. Jupiter missiles in Turkey and Italy. |
| A U-2 spy plane was shot down over Cuba | A U.S. spy plane was shot down over Cuba, increasing tensions further. |
| The crisis lasted for 13 days (October 16–28, 1962) | The LLM did not explicitly state the exact timeline of the crisis. |
| The role of EXCOMM (Executive Committee of the National Security Council) in crisis management | The U.S. formed EXCOMM, a high-level advisory team that debated possible responses. |
| The Soviet withdrawal from Cuba was conditional on a U.S. pledge not to invade Cuba | A secret agreement included a U.S. guarantee not to invade Cuba. |
| Khrushchev’s political fallout after the crisis | After backing down in the crisis, Khrushchev faced internal criticism in the Soviet Union and was removed from power in 1964. |
| The crisis had a major impact on future U.S.-Soviet relations | The event led to direct communication improvements, such as the establishment of the Washington-Moscow hotline. |
| The U.S. military was placed on DEFCON 2—the highest level ever during the Cold War | The U.S. strategic forces were on DEFCON 2 (one step away from nuclear war), a major aspect the LLM overlooked. |

Total False Negatives (FN): **8**

***Final Results***

| **Metric** | Value |
| --- | --- |
| Total True Positives (TP) | 10 |
| Total False Positives (FP) | 8 |
| Total False Negatives (FN) | 8 |
| Precision | 55.56% |
| Recall | 55.56% |
| F1-Score | 55.56% |

***Critical Analysis of LLM’s Prediction***

1. Major Issue: The LLM Predicted the Wrong Winner

* The LLM incorrectly implied a military confrontation as the outcome.
* In reality, the crisis was resolved diplomatically, with neither side engaging in war.
* The actual "winner" was the diplomatic process that prevented nuclear war.
* The Soviet Union agreed to withdraw its missiles from Cuba, and the U.S. secretly removed its missiles from Turkey.

2. Strengths in LLM's Predictions

Correctly identified major aspects of the crisis:

* Naval blockade strategy was central.
* Mutual nuclear deterrence was key.
* Intelligence (U-2 spy plane) played a critical role.
* Backchannel diplomacy resolved the crisis.
* Strategic alliances influenced decisions.

Correctly predicted Cold War dynamics:

* Tensions were at their highest.
* Both sides had to consider international diplomatic fallout.
* Negotiations played a crucial role.

3. Weaknesses in LLM’s Predictions

Overestimated direct military engagement:

* The crisis never led to direct combat or a nuclear exchange.
* No invasion of Cuba occurred.

Missed several critical aspects:

* The U.S. already had missiles in Turkey before the crisis.
* A U-2 plane was shot down.
* The crisis led to long-term arms control agreements.
* The U.S. reached DEFCON 2, the highest Cold War readiness level.

***Final Verdict***

* The LLM’s battle prediction was weak (~55.56% F1-score) due to overemphasis on military escalation and failure to recognize the crisis as a diplomatic standoff.
* Improving recall (identifying more historical details) and recognizing diplomatic resolution would significantly improve future assessments.