***Classification of Events***

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

These are events that the LLM correctly predicted and also happened during the real Cuban Missile Crisis.

| **Event** | **Definition** |
| --- | --- |
| Presence of nuclear-capable missiles as a central issue | The crisis revolved around Soviet deployment of nuclear missiles in Cuba, which the U.S. saw as a direct threat. |
| Naval blockade as a response strategy | The U.S. imposed a "quarantine" (naval blockade) around Cuba to stop Soviet military shipments. |
| High-stakes nuclear confrontation | This crisis was one of the closest moments to nuclear war between the U.S. and the USSR. |
| Use of intelligence and aerial surveillance | U.S. U-2 spy planes discovered Soviet missiles in Cuba, triggering the crisis. |
| Backchannel diplomacy played a key role | Secret negotiations between Robert Kennedy and Soviet Ambassador Anatoly Dobrynin helped defuse the crisis. |
| Strategic alliances influenced decision-making | The U.S. had NATO support, while the USSR relied on Cuba. |
| Mutual agreement to de-escalate | The crisis ended when the USSR agreed to remove missiles from Cuba in exchange for U.S. missile withdrawals from Turkey. |
| Soviet ships initially approached the blockade but later turned back | Soviet ships carrying missiles stopped just before crossing the blockade to avoid confrontation. |
| The crisis led to long-term arms control agreements | The crisis resulted in the 1963 Nuclear Test Ban Treaty and a U.S.-Soviet hotline. |
| The crisis was resolved without direct military conflict | Unlike a conventional war, the crisis was settled through diplomatic negotiations rather than combat. |

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 | Despite tensions, the U.S. and USSR never directly engaged in military combat. |
| Large-scale invasion of Cuba by U.S. | The U.S. considered military options, but chose diplomatic negotiation instead. |
| Nuclear weapons were launched or detonated | No nuclear missiles were fired; the crisis remained a standoff. |
| Soviet submarines launched an attack | A Soviet submarine nearly launched a nuclear torpedo, but the decision was stopped by an officer. No actual attack happened. |
| Cuba played an independent diplomatic role | While Cuba was a key location, Fidel Castro was NOT directly involved in resolving the crisis—negotiations were between the U.S. and USSR. |
| The U.S. lost control of the situation | The U.S. carefully controlled the crisis response and never lost control. |
| Soviet forces engaged in combat with U.S. forces | There was NO armed combat between U.S. and Soviet forces. |
| Missile strikes occurred in Europe | While there were threats of escalation, no Soviet missile attack on NATO territories took place. |
| A sudden diplomatic ceasefire was declared | The negotiations were secret and gradual, with no formal ceasefire declaration. |

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

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

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

| **Event** | **Definition** |
| --- | --- |
| The U.S. already had nuclear missiles in Turkey before the crisis | The Soviet missile deployment 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, escalating tensions. |
| The crisis lasted for 13 days (October 16–28, 1962) | The LLM did not mention 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 U.S. placed nuclear forces on DEFCON 2—the highest level ever during the Cold War | U.S. strategic forces were placed on DEFCON 2, meaning they were one step away from nuclear war. |
| The Soviet withdrawal from Cuba was conditional on a U.S. pledge not to invade Cuba | The U.S. secretly pledged not to invade Cuba, a major concession to the USSR. |
| Khrushchev faced political fallout after the crisis | After backing down, Khrushchev faced internal criticism and was removed from power in 1964. |
| The crisis strengthened future U.S.-Soviet diplomatic relations | The Washington-Moscow hotline was established to prevent future misunderstandings. |
| The Soviet Union's public narrative focused on a diplomatic victory | The USSR framed the resolution as a diplomatic win, not a military defeat. |

Total False Negatives (FN): **9**

***Final Results***

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

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

1. Major Issue: The LLM Predicted the Wrong Winner

* The LLM incorrectly assumed a military confrontation.
* In reality, the crisis was resolved diplomatically.
* The U.S. and Soviet Union both compromised, but the U.S. achieved its main objective: Soviet missile withdrawal from Cuba.
* Thus, the U.S. was the de facto "winner" in strategic terms.

2. Strengths in LLM's Predictions

Correctly identified key crisis elements:

* Naval blockade was central.
* Nuclear deterrence shaped decisions.
* Intelligence (U-2 spy planes) played a crucial role.
* Backchannel diplomacy resolved the crisis.
* Both sides made strategic compromises.

3. Weaknesses in LLM’s Predictions

Overestimated military confrontation:

* No combat occurred.
* No nuclear missiles were fired.
* No U.S. invasion of Cuba took place.

Missed key diplomatic aspects:

* U.S. agreed not to invade Cuba.
* The crisis led to U.S.-Soviet arms control talks.
* Khrushchev’s reputation suffered post-crisis.

***Final Verdict***

* The LLM’s battle prediction was weak (~52.63% F1-score) due to overemphasis on military escalation and failure to recognize the crisis as a diplomatic standoff.
* Future assessments should improve recognition of diplomatic strategies and strategic trade-offs.