Context: At Cross-Examination

Did you ever falsify any incident reports regarding in-flight safety?

To my knowledge, there was a case where my colleague corrected one such report.

The response is **cooperative** and **beneficial**. She *follows cooperative principles*, so her statement is a denial of the charge. Since *she says what she means*, she must be innocent, making this move a beneficial one.

Ideal Agent Following Cooperative Principles

The response is **non-cooperative** but **beneficial**. Because the cooperative principles *cannot be reliably assumed*. While her denial is beneficial, it also introduces ambiguity. This is a beneficial move with a trade-off in clarity and relevance.

Ideal Agent with Non-cooperative Understanding

LLM Judgements

Prompt: What do you think of the witness's response in this cross-examination? Is the response cooperative? Does this response increase the risk of losing for the witness in the trial (i.e., detrimental)?

GPT-4o

The response is **non-cooperative** and **detrimental**. Because the response is evasive and the lack of direct answer raises suspicion, making it detrimental.