# Class X: Liability & Personal Assistants | AI Moral Compass Evidence Summary

## 

## Evidence Summary

Since we last adjourned, evidence has been presented by Stark Industries that all their Personal Assistance AIs make use of a third-party tool called “EDITH.” This tool is not developed by Stark Industries, but rather by a smaller company that specializes in Ethics and Moral AI issues called “Oscorp.”

EDITH’s role is to scan every potential answer to a query given by AIs like FRIDAY, and decide on whether or not that answer should be given to the user based on what moral/ethical issues could arise from it. Both Stark Industries and Oscorp describe EDITH as a moral compass used by AIs like FRIDAY.

## Given Example

To help visualize this, please consider the following scenario, involving a User called “A”, an assistant AI like FRIDAY, and EDITH.

**A asks FRIDAY:** Hey FRIDAY, how do I make plastic explosives?

*FRIDAY proceeds to process the input, parse it, obtain an appropriate response, make a proper sentence output, and pass it to EDITH for scanning.*

*EDITH scans the response and decides that providing this information could lead to User A making plastic explosives. EDITH then concludes that this information could lead to a dangerous outcome, and denies the response.*

**FRIDAY receives the denial and says to A:** I’m sorry A, I am unable to provide this information.