What more could I have done?

Stress test it with multiple conversation flows to find points of failure.

Identify points of failure and best practices to handle it gracefully.

Giving JARVIS a personality.

"Go get yourself some coffee!"

Did I Evaluate it?

Due to time and implementation constraints, I couldn't really test it. I would have manually tested it by observing meetings for another 2 weeks to note the time spent by using the new system and conducted a second round of interviews.

Followed by a third...a fourth...