Assignment 1 - Short Definitions

1. Proactive Triage

Proactive triage is the capability of an agent to self-assess and categorize incoming data preceding a user action. Rather than waiting for a response, the agent will assess and determine an incoming piece of information, for instance, an email, if it needs to be disregarded, highlighted for human review, or taken action on.

2. Persistent Memory

Persistent memory allows the AI agent to maintain recall capabilities between interactions. This incorporates elements such as user preference, decisions, corrections, and feedback. Through recall, the agent meets the required adaptability standards, as it does not require the creation of new memories every time it undertakes a different task.

3. True Autonomy via Human-in-the-Loop (HITL)

True autonomy in Human-in-the-Loop is a design principle wherein the agent pursues autonomous execution for low-risk operations but seeks human validation for high-risk operations. In this case, the human becomes a safety net that can approve, reject, or modify the agent's decisions.

4. Robustness via Evaluation

Robustness by evaluation ensures the working of the agent in a reliable manner under change of conditions. It includes the testing of the agent using pre-defined data sets, measuring accuracy, and evaluation of the quality of the response that comes with techniques such as LLM-as-a-judge. Continuous evaluation allows the detection of weaknesses, enhances decision-making, and maintains system reliability.

5. Real-World Deployment

Real-world deployment involves relocating an AI agent from a prototype environment to a live production environment. The process includes integrating with live services such as email and calendar APIs; handling users' data securely, and ensuring the system works effectively and safely under practical applications.

6. Developing a User Interface

Building a UI requires the creation of an interface where the users are able to interact with the agent. This interface enables the users to view decisions, approve or edit them, as well as receive notifications and comprehend actions performed by the agent. Building an effective UI enhances usability and the overall interface.