**QnA**

1. **Memory and context:**

Wang explained that while models like GPT excel at short-term contextual reasoning, LinkedIn’s H Assistant had to solve a harder problem - persistent conversational and experiential memory across sessions and features. Their architecture stores context as structured “sessions”/“episodes” of user-agent interactions and retrieves them using both vector embeddings and symbolic keys, allowing a member’s history or recruiter’s past preferences to inform later recommendations.

For the challenges, he highlighted: (a) maintaining privacy and access control when memories span products, and (b) ensuring the retrieval system scales without amplifying outdated or biased context.

1. **Future directions:**

Wang mentioned several upcoming use cases: how to help member learn deeper and more insights about companions through tools that summarize long-form professional exchanges (e.g., interview transcripts or collaborative article threads). He emphasized compound AI agents orchestrated by a central coordinator as LinkedIn’s next frontier. These will extend personalization and support across the entire member journey while preserving reliability and human oversight.

1. **Open Thoughts**

I was impressed by how mature LinkedIn’s platform already is, yet I’m still curious about how they evaluate “trust” and “memory freshness.” Persistent context can be powerful but also risky if the stored data reflects outdated user intent. It would be interesting to learn how the team audits long-term memory for accuracy and fairness while maintaining scalability.