EVA Limitations

There are some limitations to EVA. We'll take a look at those.

💬 Sensitive to Prompt

Sensitive to Prompt means the quality of the response can be highly dependent on the prompt given.

> ℹ️ Workaround: **Prompt Tuning** – If we use keywords, terminologies, and jargons same as source the quality of response can be improved.

For example, if we ask EVA on test cases for Recommendation end slate. EVA cannot recognize any test cases as the Recommendation End Slate is referred as Player Recommendation in Test Rail. If we revise the query to check existing test cases for Player Recommendation, EVA will be able to find the correct details for the query.

🗑️ Obsolete Data in Source

Second challenge is Obsolete or Outdated Data in source systems. The quality of the source data directly impacts the results. Clean and accurate data in source systems or identifying and implementing any logic or approaches to clean data during EVA data processing can also help to improve response. Filtering and cleansing in EVA processing would may require some guidance, please share your input/feedback with Unified QE team.

> ℹ️ Workaround: **Targeted Query on Latest Data** – Similar to previous case, if we revise the prompt in include keywords from latest page details we can get appropriate response. Also, we can exclude outdated data upon request.

For example, there are different pages with details on Webplex Connected TV team. Some of these are archived. Also, there is a latest page on active team members. If you look at QA team members, there is a change. If we will directly query for Webplex Connected TV team. EVA responded with details; however, this information is outdated. As a workaround, we may add some filtering logic like excluding pages under paths with keyword archived in the EVA processing (We can implement such logic in EVA processing upon confirmation). Alternatively, we can revise prompt to add specific keywords like Network Streaming Scrum team, to point EVA to right page.

—

↗ Context Size Limitation

Generic questions that require a large amount of data to be analyzed may not yield accurate results, unless there are some consolidated statuses or references are out there.

> ℹ Workaround: **Targeted query on Smaller Sets** – The application performs better when asked targeted questions on specific content.

For Example, Under the playback section, we are having around 64 test cases. However, we have a limited context of 15 items per query. If we ask EVA on Regression cases for Playback. Due to limited context, it could only identify 5 test cases. If we revise and target our query to specific section like Ads UI, which has total 6 test cases consist of 1 smoke and 5 regressions, EVA will be able to provide accurate response as 5 regression test cases, and one smoke test.