**Problem Statement:** explore how AI tools (like ChatGPT, Claude, or Microsoft Copilot) can help generate valid and invalid test data for a simple history store functionality of an OTT app and document the findings.

**Scope**

1. Test Area: 1. Before login 2. After login
2. Data Coverage:
   1. Invalid data
   2. Valid Data
   3. Edge case Data
3. Tools to Explore: ChatGpt, Teams Copilot, Claude, Browserstack Agents

**Use same Prompt and compare result**

**Prompt\_1**: Act as a QA and generate a Valid test data in Table format to test a simple history store functionality of OTT app

**Prompt\_2**: Act as a QA and generate 2 negative/invalid test data in Table format to test a simple history store functionality of OTT app

**Prompt\_3:** Suggest additional fields that should be tested in OTT watch history (e.g., DeviceType, GeoLocation) and generate valid+invalid test data for them.

**Chat GPT(Think Longer):**

Observation:

1. It created the valid test data as per the Prompt
2. Created test data includes the positive flows cases and we can use this

**Created Test Data:**

**Prompt\_1-** [**https://docs.google.com/spreadsheets/d/15J5gH-8CAlq8DwaA02jIbxHUDH6Jk9NW2gXvKOPF9zc/edit?gid=1056169268#gid=1056169268**](https://docs.google.com/spreadsheets/d/15J5gH-8CAlq8DwaA02jIbxHUDH6Jk9NW2gXvKOPF9zc/edit?gid=1056169268#gid=1056169268)

**Prompt\_2:** [**https://docs.google.com/spreadsheets/d/15J5gH-8CAlq8DwaA02jIbxHUDH6Jk9NW2gXvKOPF9zc/edit?gid=1293528662#gid=1293528662**](https://docs.google.com/spreadsheets/d/15J5gH-8CAlq8DwaA02jIbxHUDH6Jk9NW2gXvKOPF9zc/edit?gid=1293528662#gid=1293528662)

**Prompt\_3:**

[**https://docs.google.com/spreadsheets/d/15J5gH-8CAlq8DwaA02jIbxHUDH6Jk9NW2gXvKOPF9zc/edit?gid=1079019117#gid=1079019117**](https://docs.google.com/spreadsheets/d/15J5gH-8CAlq8DwaA02jIbxHUDH6Jk9NW2gXvKOPF9zc/edit?gid=1079019117#gid=1079019117)

**Claude(Sonnet 4):**

**Observation:**

1. It created the all type test data which is not even added in prompt- valid, invalid, edge,Retrieval Testing, Management Operations,Concurrent Sessions.
2. Test data covers various factors- Multiple user profiles, Different content types, Various devices, Different video qualities, Multiple genres and watch progress percentages,Realistic timestamps and durations
3. Easy to understand as it added notes and also have given the summary which includes high level information

**Created Test Data:**

**Prompt\_1:** [**https://docs.google.com/spreadsheets/d/15J5gH-8CAlq8DwaA02jIbxHUDH6Jk9NW2gXvKOPF9zc/edit?gid=0#gid=0**](https://docs.google.com/spreadsheets/d/15J5gH-8CAlq8DwaA02jIbxHUDH6Jk9NW2gXvKOPF9zc/edit?gid=0#gid=0)

**Prompt\_2:** [**https://docs.google.com/spreadsheets/d/15J5gH-8CAlq8DwaA02jIbxHUDH6Jk9NW2gXvKOPF9zc/edit?gid=1371978441#gid=1371978441**](https://docs.google.com/spreadsheets/d/15J5gH-8CAlq8DwaA02jIbxHUDH6Jk9NW2gXvKOPF9zc/edit?gid=1371978441#gid=1371978441)

**Prompt 3:** <https://docs.google.com/spreadsheets/d/15J5gH-8CAlq8DwaA02jIbxHUDH6Jk9NW2gXvKOPF9zc/edit?gid=1234347097#gid=1234347097>

**Comparison Points**

1. Differences:

1. ChatGpt generated the test data as per the ask. Which I mentioned in the prompt accordingly provided the data whereas Claude generated all kinds of test data irrespective of what i ask. It think further and generated all possible cases test data

2. Claude seems more creative

1. Similarity

1.Table format is similar, both generated the test data with same details like- ID, user profile, content, watch history details, devices, duration, expected result

2. Covers All scenario’s/cases test data even ChatGpt just all possible Valid data with different variety of cases.

1. Negative

1. The Test data they created may be we can’t be able to as it is as it gives the structure and ideas about what kind of data we can use.

2. On actual workflow we might need to review it before use as some test data may be not available or may be not more relevant to your case

1. Positive

1. AI is a model which is trained on multiple datasets so it can think more smartly than manual QA and it will give us ideas like how many or more test cases we can cover with a variety of test data.

**AI reduced your effort vs. if you did this manually.**

1. So of course it seems helpful for quickly generation the test data and also covering most of the scenario’s
2. In Manual we may require more time to go through test cases and then based on that find out the resources from where we can get the test data then manually by adding query to SQL or may be manually one by one finding test data.  
   So with AI all these efforts are getting reduced and we are able get the not exact but more synthetic,variety and, diverse test data
3. Even if Test data is generated in less time, Manual review is required as we can’t trust 100% of the machine before using any data. As may be the created data may not be relevant or may be missing something so we might need to create our own test data by referring it as an example.