## User-centered Validation Session Protocol

**1. Participant Selection**

Participants are selected based on their working experience in issue tracking systems and the “*Non-profit organization care” (NPO)* repository. Participants are invited by e-mail, which outlines the study’s and user experience (UX) testing session’s objectives and an overview of the session.

Participants will be provided with an informed consent form (see User-centered validation session – Informed consent.docx). If participants give consent, we proceed with the UX testing session.

**2. User-centered Validation Session Preparation**

The GDI system runs on a desktop with a local Neo4j Instance with the knowledge graph of NPO and large language model (LLM), specifically Llama 3.1. Additionally, NPO is opened in Azure DevOps web browser, and a Notepad++ file is created and opened for the participant.

**2. Overview**

Before starting, participants will be asked if they have any questions or need further clarification.

1. **Brief Introduction (15 minutes)**

* The GDI approach uses knowledge graphs, retrieval-augmented generation, and LLMs to enable users to query work items in natural language and trace how the LLM integrates the user query with the retrieved information to generate a response.
* The study aims to explore user experiences with the system in supporting software development teams in retrieving information to address their information needs.
* We provide a short overview of the system’s browser-based interface and functionality. Moreover, we present the knowledge graph of NPO and how the participants can navigate through the graph and execute queries in Neo4j.

1. **Hands-on Testing (30 minutes)**

* We start the recording of the Microsoft Teams meeting and the screen of the desktop is shared. Participants are given control of the desktop where GDI is running.
* We defined five user questions with varying complexity related to TLR (see UQ1 – Q5). The questions and guidelines were provided as a cheat sheet to help participants, who were encouraged to modify the questions to suit their needs.
  + UQ1: Is there a relationship between Bug ID and User\_Story ID?
  + UQ2: What are the tasks of User\_Story ID?
  + UQ3: What is the longest dependency chain in the system?
  + UQ4: Which is the longest dependency chain in the system?
  + UQ5: Which User\_Stories are most similar based on their title?
* For each question, they will search for the answer in Azure DevOps and write down their expected answer in Notepad++. The participants will interact with the system to retrieve information and compare their answer to their expected answer.
* They will be asked to think aloud while doing these actions.
* During the hands-on testing section, the researchers will not interfere unless participants ask for guidance, which will be reported in the paper.

1. **Interview (15 min):**

* Wediscuss the experience of the participant, using questions to validate the usability, accuracy, adequacy of the retrieved context, usefulness, and likeliness of adopting the GDI system (see Questions.docx).
* Moreover, we aim to gather feedback to refine the GDI system, improve retrieval accuracy, and enhance user experience.

1. **Closing:** We thank the participant for their time.