**Abstract for Artifact Evaluation**

**Paper title**: Benchmarking Automated GUI Testing for Android against Real-world Bugs

**Purpose of the artifact**: Our artifact, Themis, offers the first ground-truth empirical evaluation of automated GUI testing for Android, and gives the insights/answers to an overarching question "How effectively and thoroughly can these tools find crash bugs in practice?". Themis includes (1) a carefully constructed dataset with 52 real, reproducible crash bugs, and (2) a unified, extensible infrastructure with six recent state-of-the-art testing tools.

**Badge we are claiming**: Functional and Reusable

**Required technical skills**: The familiarity with Android is preferable.

**Artifact evaluation requirement**:

1. We provide a VM image with VirtualBox for evaluation. Your computer needs at least 16G of memory, and at least 40G of storage. We do not have specific requirement on the operating system.

2. Because Themis benchmarks automated GUI testing tools for Android. The artifact evaluation may require a long evaluation time. We suggest to run one testing tools

In our own evaluation, we allocated 6 hours for each testing tool to run on each bug and repeated this process for five times.