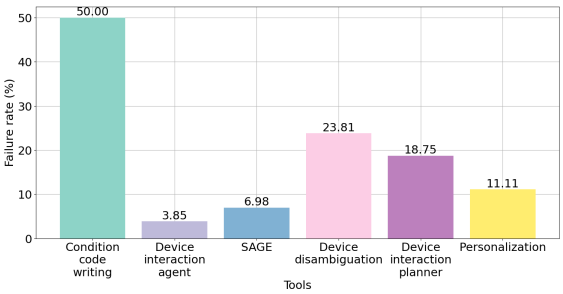

Figure 9: Failure rates for different components of the system. Failure rate is computed as the ratio of the number of tests which failed due to a failure in the given component and the total number of tests using the component. Failures are attributed to components by manual analysis.

be grounded with respect to user preferences, devices, and the external world, we devised and presented several novel tools. We also created a dataset of challenging smart home automation test cases which tested the system's ability to be personalized, to resolve users intents from unstructured queries, to resolve devices referred to in natural ways, and to command persistence and chaining. achieved a success rate of around 51% on these tasks. This value, while far from perfect, is much higher than that achieved by existing LLM-based home automation solutions. Each success required the successful sequential application of many tools, so in fact the number of successful tool uses is much larger than the number of failed ones. As such, SAGE represents a highly promising first step towards the creation of truly flexible smart home agents that users can interact with as naturally as they would with a close friend.

9 Appendix