



# Alan's Speakeasy - An Ecosystem for the Evaluation of Conversational Agents

Oana Inel (\*), Luca Rossetto(\*), Cristina Sarasua (\*), and Abraham Bernstein (\*), equal contribution  
Dynamic and Distributed Information Systems Group, Department of Informatics

## Idea

With the increasing interest in conversational agents, it became relevant to have a framework to evaluate and compare conversational agents (aka bots). We implemented an ecosystem including (1) a Web-based platform that hosts chatrooms where humans and conversational agents can interact for humans to evaluate the agents in a Turing-test style, and (2) a set of artificial agents whose goal is to help in the evaluation.

## Speakeasy Capabilities

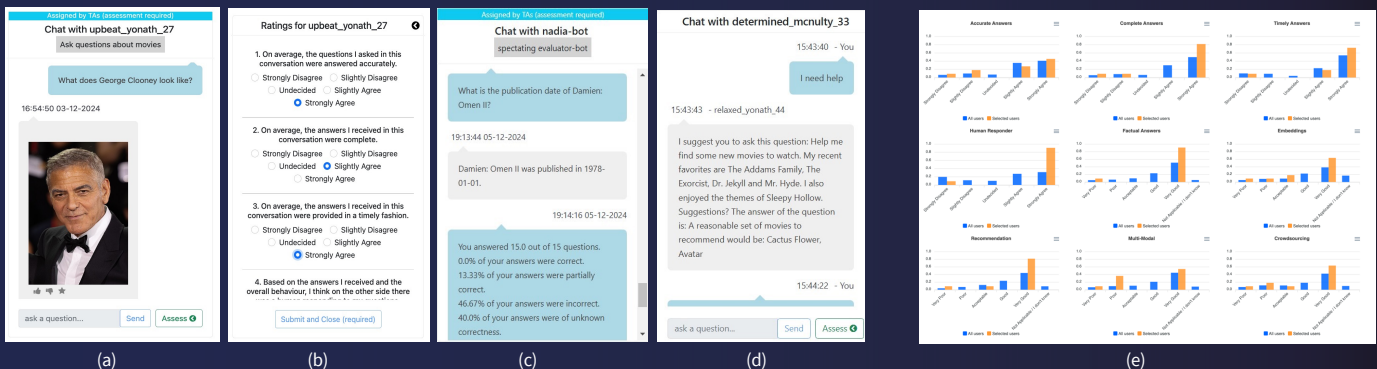
- Generation of conversational sessions by pairing users (i.e., humans and bots) and configuration of conversations (e.g., conversation duration, evaluation dimensions).
- Human feedback provision at the level of a conversational agent message (i.e., like/dislike/star) and entire conversation (i.e., through a customizable survey containing Likert-scale questions).
- Automated feedback provision through Assessor Bots. Assessor Bots can 1) assist human evaluators in their assessment of other agents, 2) conduct evaluations autonomously, and 3) assist students with their conversational agents' development.
- Customizable data generation pipeline for preparing suitable datasets based on Wikidata and other movie-related sources.

## Innovative because...

- We made AI technologies accessible through a hands-on project-based interactive learning experience.
- Speakeasy provides automated evaluation support.
- Speakeasy can host multimodal conversations.

## Exemplary because...

- It is open-source, and it is based on established industry standards.
- It is configurable and extensible.
- Welcomes multi-disciplinary scientific and educational work on conversational agents.
- Any conversational agent can be connected via the Speakeasy API.



## Success Factors / Results

- We deployed the Speakeasy ecosystem in four iterations of the Advanced Topics in Artificial Intelligence (ATAI) course (2021 - 2024) and we plan to continue using it in forthcoming years.
- In total, 337 people (students, lecturers, and researchers) participated in the evaluation of 114 conversational agents.
- We conducted interviews with eight experts, resulting in 34 new requirements for improving and extending Speakeasy's purpose in teaching and research.
- Speakeasy Demo paper accepted at the TheWebConf 2025.



A project supported by the funding line open\_innovation

Contact: [speakeasy-info@ifi.uzh.ch](mailto:speakeasy-info@ifi.uzh.ch)