# A Mirror to Human Question Asking: Analyzing the Akinator Online Question Game



## Introduction

Question asking is a critical aspect of human communication. Yet, little is known about the reasons that lead people to ask questions; which questions are considered better than others; or what cognitive mechanisms allow the ability to ask informative questions.

Here, we take an empirical step towards investigating human question asking. We do so by an exploratory data-driven analysis of the questions asked by Akinator, a popular online game of a genie who asks questions to guess the character that the user is thinking of, as a reflection of how humans might ask questions.

## **Materials & Methods**

## **The Akinator**

An online game where a magical genie – the Akinator - asks users to think of a character (real or fictional), object or animal (). Akinator will ask the user "yes" or "no" questions until it guesses who or what the user has thought of.



#### **Question Collection & Analysis**

Questions asked by the Akinator – 7,473 distinct questions were collected by playing 4,000 games in two runs.

Questions by games – 1,500 games were played with constant answers throughout the entire game.

**First questions** – 10,000 games were initiated in two different runs and the first question of each game was recorded.

**Real games on specific characters –** multiple games were played about specific known characters (e.g., Mickey Mouse, Mark Zuckerberg).

#### **Topic Modelling of the Questions**

LDA Topic Modelling – Latent Dirichlet Analysis (LDA) is the most frequently used method for performing topic-modelling and is based on probabilistic assumptions. We used the model with the questions dataset, having each question treated as a document.

**BERT Topic Modelling** – Based on a continuous representation of topics by jointly embedding word, document and topic vectors such that the distance between the vectors represents the semantic similarity between them. We use this model on the dataset of questions, in order to score the words relevant to each topic.

#### Gal Sasson and Yoed N. Kenett

Technion – Israel Institute of Technology

#### Results



**Figure 1.** The number of new distinct questions found through 300 games played with a constant answer ('yes', 'no' or 'probably'). This number consistently decreases as more games are played, until only a few new questions are found with every new game.



Figure 2. The top 20 most frequent questions in 4000 games. The games were played with constant answers throughout the whole game, so that each of the five possible answers ('yes', 'no', 'i don't know', 'probably' and 'probably not') was used in 800 of the games.

Question	Run 1 – November 2021	Run 2 – December 2021	Total
Is your character a girl?	1132	2164	3296
Is your character a Youtuber?	1603	8	1611
Does your character make Youtube videos?	1115	0	1115
Is your character known for making Youtube videos?	1007	0	1007
Is your character a woman?	143	42	185
Is your character real?	0	268	268
Is your character a real person?	0	413	413
Does your character know your name?	0	337	337
Is your character a female?	0	920	920
Does your character personally know you?	0	651	651
Have you ever talked to your character?	0	8	8
Is your character a make?	0	189	189
	5000	5000	10000

Table 1. Questions that were asked as first questions in games and the number of times they appeared across two separate runs. Clearly, the set of questions changed over time.





COMPLEXITY LABORATORY



## **LDA Topic Modelling Results**











#### Figure 3.

The 20 highest weighted words in each topic created by the LDA model.

#### **BERT Topic Modelling Results**















Topic #11

make

eos









Figure 4. The 20 highest weighted words in each of the largest topics created by the BERT model. In total, 44 topics were created.

### **Conclusions**

video

Our results provide novel insights into the process of asking questions, via an online question asking game. We find that this process encompasses multiple question asking strategies, likely mimicking human question asking.

Our effort is significantly constrained by patent constraints of the software. Yet, it illustrates the potential of analyzing such online question asking games via text-based methods, deriving further insight into the time unfolding process of asking questions.









