

# Can You Distinguish Truthful from Fake Reviews?

## User Analysis and Assistance Tool for Fake Review Detection

Jeonghwan Kim\*, Junmo Kang\*, Suwon Shin\* and Sung-Hyon Myaeng  
 Korea Advanced Institute of Science and Technology, School of Computing

\* is for equal contribution

### Overview

- Customer review authenticity has become crucial in e-commerce platforms.
- We propose a simple assistance tool, **You Only Need Gold (YONG)**, to detect deceptive reviews and augment user discretion.
- We provide an in-depth **user understanding on dealing with fake reviews** under the guidance of YONG.

### You Only Need Gold (YONG)

#### Fake Review Detection Tool

Write review down in the text area and control the pe.

The Chicago Palmer House disappointed me. I chose it because it a Hilton hotel, but it definitely does not live up to the standard set by this name. My first impression was a negative one when I was greeted by a rather rude staff. My room might have been luxurious years ago, but the general impression was that it had not been cleaned regularly and could use a little freshening. The bathroom was obviously newer but the shower was difficult to use and the water pressure was absolutely horrible. The food served by the room service and at the hotel's restaurant sounded delicious on the carefully crafted menus, and the high prices made me expect delicate cuisine. The food simply did not deliver its promises. Everything tasted very bland and the portions were ridiculously small. The location of the hotel might seem like an advantage, since it is in downtown Chicago nearby many touristic attractions; however, the windows in my room did not protect me from the constant noise of a luxury hotel and did not use the hotel name and price tags.

99% Fake

the chicago palmer house disappointed me. I chose it because it a hilton hotel, but it definitely does not live up to the standard set by this name. my first impression was a negative one when I was greeted by a rather rude staff. my room might have been luxurious years ago, but the general impression was that it had not been cleaned regularly and could use a little freshening. the bathroom was obviously newer but the shower was difficult to use and the water pressure was absolutely horrible. the food served by the room service and at the hotel's restaurant sounded delicious on the carefully crafted menus, and the high prices made me expect delicate cuisine. the food simply did not deliver its promises. everything tasted very bland and the portions were ridiculously small. the location of the hotel might seem like an advantage, since it is in downtown Chicago nearby many touristic attractions; however, the windows in my room did not protect me from the constant noise of cars going by. overall, this hotel would be decent if it did not advertise itself as a luxury hotel and did not use the hotel name and price tags.

- YONG provides the **gold indicator**, which consists of three intuitive, distinct features:
  - Model decision** (Fake / Gold)
  - Probability (%)**
  - Evidence** (word highlights – the more intense the color highlight, the more important of a role the word plays)
- YONG uses **BERT** as the backbone model, which is finetuned on the OpSpam dataset, with its attention weights visualized as Evidence.
- We leverage YONG to run user evaluations and on human capabilities and tendencies in detecting deceptive reviews.

### Research Questions (RQs) / Experiment

- Can humans **outperform** models in fake review detection?
- Can YONG **augment** human capability?
- Which feature in YONG influences human decision the most?

- Separate **experiment for each RQ**
- 24 participants are required to classify fake reviews with & without YONG.
- The test is **single-blind**; participants don't know the ground-truth label

#### Experiment #2

Detect Fake Review with Machine indicator

Review #2 98% Gold

Very disappointed in our stay in Chicago Monoco. We have stayed many times elsewhere, primarily in Washington DC and are accustomed to great customer service, beverages like water or soda at the wine bar, coffee and papers in the morning, help with bags. Not only did the Chicago monoco do none of these things, the staff was not helpful, either. Requests were not honored and the staff did not seem happy to be there. You got the feeling you were 'bothering' people if you asked a question. No bellman, the doorman did not open the door or help with bags. Even though we were traveling with a child, I had to request a fish when they did not bring it. Really baffling. **we love the Monoco in Washington.**

Fake Gold

Gold is selected. Is it right?

### Results

Condition	Score	Feature	Decision	Probability	Evidence
No tool	0.41	Influence	3.69	3.91	1.87
With tool	0.54				
Model	0.70				

Results of Experiments 1-2

Results of Feature-wise Influence (1-5 scale)

- Humans are not good at detecting fake reviews, underperforming the model by a large margin.
- With YONG, the accuracy increases substantially. **(0.41 → 0.54)**
- Among three features of YONG, **probability** plays the primary role in convincing users.

### Discussion

- Human capability of fake review detection is **unreliable** and requires machine assistance.
- Evidence (interpretable attention visualization) is **hardly explicable**.
- Interpretability** is different from **explainability**.
- Assistive tools need to provide **faith-gaining features**.