

# Interactive Dialogue Interface for Personalized News Article Comprehension

**Tomoya Higuchi**

The University of  
Electro-Communications  
1-5-1, Chofugaoka, Chofu,  
Tokyo, Japan  
h2430109@edu.cc.uec.ac.jp

**Michimasa Inaba**

The University of  
Electro-Communications  
1-5-1, Chofugaoka, Chofu,  
Tokyo, Japan  
m-inaba@uec.ac.jp

## Abstract

We developed an interface to explain news articles through dialogue by considering the user's comprehension level. The interface generates several pertinent questions based on the ongoing dialogue and news article, and users advance the conversation by selecting a question. Based on the user's selected questions, the interface estimates their comprehension level of the news article and adjusts the difficulty of the generated questions accordingly. This enables a personalized dialogue tailored to each user's comprehension needs. The results of the baseline comparison experiments confirmed the usefulness of the interface.

## 1 Introduction

Organization for Economic Co-operation and Development reported that 21% of individuals in the United States aged 16 years and over have low English literacy. This limitation results in them having difficulty reading longer texts such as news articles (Desjardins et al., 2013). In this paper, we construct an interactive news commentary dialogue interface that considers the user's comprehension level to increase the understandability of news articles.

We constructed an interface, shown in Figure 1, that interactively explains news articles. The interface generates several pertinent questions based on the ongoing dialogue and news article, and users advance the conversation by selecting a question. This interactive approach is expected to provide users the opportunity to organize information and deepen their understanding. When users select a question, the system predicts their comprehension level of the news article based on the selected question and adjusts the difficulty of the next question accordingly. This feature allows the dialogue to be tailored to each user's knowledge level.

The screenshot displays a dialogue interface with the following content:

- Title:** Brightest-ever cosmic explosion solved but new mysteries sparked
- System:** Researchers have identified the source of the brightest burst of light ever recorded, which originated from an exploding star observed in 2022.
- User:** What does it mean when a star explodes?
- System:** When a star explodes, it undergoes a supernova, a powerful and luminous stellar explosion. This event is significant because it is thought to be a key source of production for heavy elements in the universe, such as gold and platinum, although the recent findings challenge this understanding as no heavy elements were detected in the observed explosion.
- Candidate questions:**
  - Level 1 question: What are heavy elements, and why are they important?
  - Level 2 question: How does the recent finding of no heavy elements in the supernova challenge the existing theories about the origin of elements like gold?
  - Level 2 question: What are the implications of this discovery for our understanding of how precious metals are formed in the universe?

Figure 1: Proposed Dialogue Interface. The dialogue history, which includes the user-selected questions and the system's answers, is displayed at the top. The user selects a question from the candidate questions at the bottom. The difficulty level of the displayed question candidates varies from turn to turn based on the estimated user comprehension level.

## 2 Related Works

Media often use question-and-answer series about news stories because of their clarity, but reporters manually create them; thus, the process is inefficient due to high costs and time requirements. Laban et al. (2020) proposed a method to automatically build chatbots from news articles. In these chatbots, the system presents the user with candidate questions. Podcasts with automatic construction containing QA sessions on news were also analyzed, confirming the effectiveness of a conversational format (Laban et al., 2022). However, these questions are not customized to individual user characteristics, and the tailored dialogues are not generated for individual users. User characteristics, including their social group, influence the types of questions asked. Stewart and Mihalcea (2022) introduced a method for producing ques-

tions that consider these characteristics, training a text generation model on social media data that accounts for social groups and their expertise in specific domains. An et al. (2021) developed a conversational agent prototype that incorporates a strategy to account for the user’s knowledge and ignorance regarding speech, validating the importance of considering user knowledge. Drawing inspiration from these methodologies, we generate questions that incorporate the user’s comprehension level of the news story.

### 3 System Overview

We constructed an interactive interface to explain the contents of news articles, illustrated in Figure 1. The interface first presents the user with a brief introduction to the news article and three candidate questions. The user then selects one of these questions, and the system explains the article’s content by answering the selected question. This approach reduces the user’s burden by eliminating the need to think about questions independently. The purpose of this interface is to facilitate understanding by providing explanations tailored to individual users. Candidate questions should reflect those the user might want to ask, considering the user’s comprehension level. Therefore, we generate and present candidate questions, considering the user’s comprehension level. The system comprises three modules, as shown in Figure 2: an introduction generation module, a candidate question generation module, and an answer generation module. Each module uses GPT-4 (OpenAI et al., 2024) as the large language model.

#### 3.1 Introduction Generation

The introduction generation module create an introductory summary of the news article intended for the user. This process involves utilizing the news article’s content and providing guidelines within the prompt to craft a concise introduction in one sentence. GPT-4 generates this introduction in a zero-shot.

#### 3.2 Question Candidate Generation

The question candidate generation module creates a set of questions to present to the user. A dialogue tailored to individual users must pose questions that match the user’s comprehension level. Additionally, the questions should follow the dialogue naturally and be answerable based on the news. We use GPT-4 to overgenerate candidate questions that

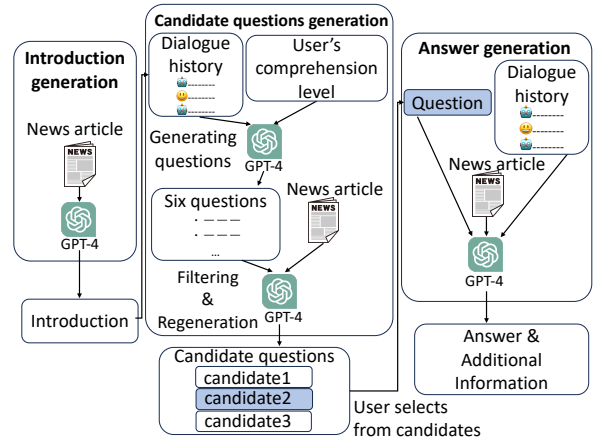


Figure 2: Architecture of Proposed system. The system comprises three modules: an introduction generation module that creates an introduction to the news article, a candidate question generation module that generates three candidate questions to present to the user, and an answer generation module that generates answers to the user’s chosen questions.

Table 1: The user’s comprehension level  $x$  and the number of candidate questions to present for each level

$x$	level 1	level 2	level 3
$x < 1.5$	1	2	0
$1.5 \leq x < 2$	1	1	1
$2 \leq x < 2.5$	0	2	1
$2.5 \leq x$	0	1	2

consider the user’s comprehension level from the dialogue history, filter them based on their answerability from the articles, and regenerate them if the required number is not fulfilled, and present three to the user.

#### 3.2.1 Candidate Question Generation

We categorize the levels of difficulty for the questions as 1, 2, and 3. Level 1 pertains to queries regarding the interpretation of terms, tailored for novices in the field; level 2 contains general inquiries about the article’s content; and level 3 comprises more intricate questions designed for field experts. The quantity of questions presented to the user for each difficulty level is modified based on Table 1, according to the user’s comprehension level estimated for each turn.

The average difficulty level of questions selected by the user up to the current turn is used to estimate the user’s comprehension level. Because the user’s comprehension level cannot be estimated in the first turn, candidate questions are presented to the user at difficulty levels 1, 2, and 3.

The difficulty level of the candidate questions is determined by referring to Table 1. The dialogue history and the difficulty condition of the candidate question are input into GPT-4, which generates a total of six candidate questions at difficulty levels 1-3, two for each level, that naturally follow the dialogue. Generating questions at levels 1 to 3 in every turn increases the likelihood that questions appropriate to each level are generated.

### 3.2.2 Filtering & Regeneration

When GPT-4 generated questions with a news article as input, the questions often anticipate information from the article not included in the dialogue history. This can prevent users from organizing information. Therefore, candidate questions are first overgenerated solely from the dialogue history, and only those questions that can be answered using the content of the news article are selected through filtering. If the number of candidate questions after filtering is less than required, the candidate questions are regenerated.

Of the six candidate questions generated, level 2 and level 3 candidate questions and the text of the news article are input into GPT-4, which determines whether they can be answered from the article’s content and filters them accordingly. No filtering for level 1 questions because they are designed to help users obtain prerequisite knowledge not in the content of the news article.

If the number of questions that can be answered from the content of a news article at each difficulty level is less than the number specified based on Table 1, GPT-4 generates the missing candidate questions by using the text of the news article and the dialogue history as input.

### 3.3 Answer Generation

The answer generation module generates responses to user-selected questions from candidates produced by the question candidate generation module. Inputs include the news article text, dialogue history, and questions, with GPT-4 generating answers and supplementary information to transition to the next topic in one shot. This approach helps prevent subsequent question candidates from focusing solely on one topic by providing context beyond just the answers.

## 4 Experiment

To evaluate the usefulness of this interface in reading news articles, we conducted a subject exper-

iment to compare it with three different baseline settings. Crowdfunder read a randomly assigned article in a randomly assigned setting and completed a comprehension test and questionnaire.

### 4.1 Comparative Methods

As comparison methods, we conducted experiments using the following three settings:

#### 1. Reading news article

We adopted this baseline to compare interactive and non-interactive formats. The user reads a news article using a web browser.

#### 2. Microsoft Copilot<sup>1</sup>

We adopted this baseline to examine the effectiveness of presenting users with candidate questions. The user opens the news article in Microsoft Edge<sup>2</sup> and asks the Copilot a question without reading the body of the news article.

#### 3. W/o comprehension level

This baseline is compared with the proposed method to investigate the effectiveness of presenting comprehension-aware candidate questions. In this baseline interface, the system presents candidate questions at a single difficulty level to the user without considering their comprehension level. The difficulty level of the candidate questions corresponds to Level 2 in the proposed method.

### 4.2 Experimental Setup

We recruited 80 crowdworkers using Amazon Mechanical Turk<sup>3</sup> and conducted a subject experiment. Four news articles were selected from BBC news, one each from the fields of natural science<sup>4</sup>, technology<sup>5</sup>, international<sup>6</sup>, and finance/economics<sup>7</sup>. Each article was chosen based on the criteria that it was at least 500 words long and contained specialized content. Each crowdworker was randomly assigned one of four settings and one of the four news articles. After reading the article, the crowdworkers took a comprehension test on the content of the news article and completed a questionnaire.

<sup>1</sup><https://www.microsoft.com/ja-jp/microsoft-copilot>

<sup>2</sup><https://www.microsoft.com/ja-jp/edge?form=MA13FJ>

<sup>3</sup><https://www.mturk.com/>

<sup>4</sup><https://www.bbc.com/news/science-environment-68787534>

<sup>5</sup><https://www.bbc.com/news/business-68225115>

<sup>6</sup><https://www.bbc.com/news/world-us-canada-68883659>

<sup>7</sup><https://www.bbc.com/news/world-europe-68761491>

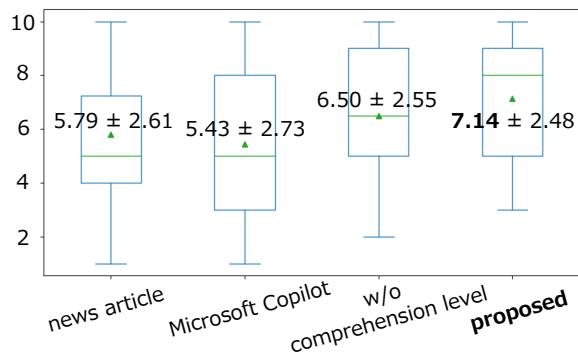


Figure 3: Boxplot of comprehension test scores.  $\triangle$  is the average score.

The comprehension test consisted of 10 four-choice questions, with one point per question. We collected a total of 80 data, five for each combination of setting and article.

### 4.3 Result

Figure 3 shows the aggregate results of the comprehension test. The average score was highest for the proposed interface. We can conclude that the proposed method, which presents candidate questions considering the comprehension level, is useful for reading comprehension. There are statistically significant differences in average comprehension test scores between the proposed interface and Microsoft Copilot ( $p < 0.05$ ) and no significant differences in the other combinations. Figure 3 shows that in the proposed interface, few people scored low on the comprehension test. This finding indicates that candidate questions considering the user’s comprehension level are effective for users with low comprehension.

According to the open-ended questionnaire, participants found w/o comprehension level interface and proposed interface to be fun and easy (e.g., "Fun to read, easy to use"). In the proposed interface, participants expressed favorable opinions about the candidate questions that considered comprehension level (e.g., "Great range of follow-up questions," "This will be sufficient for first-time readers"). However, some participants complained about the response time (e.g., "Was a bit slow to generate responses.")

## 5 Conclusion

We developed a news commentary dialogue interface that considers the user’s comprehension level. The interface alleviates the user’s burden by pre-

senting multiple automatically generated question candidates and having the system answer the selected questions. It aims to tailor the dialogue to each user by predicting their comprehension level of the news content on the basis of the selected questions and presenting question candidates that account for this level of understanding. In our experiment, we implemented the proposed interface and quantitatively evaluated its effectiveness by using crowd-sourcing. Comparison experiments between the proposed interface and baselines confirmed that the proposed interface enhances users’ reading comprehension.

In the open-ended responses to the questionnaire during the experiment, one participant said, "Conveying the content of news solely through text is difficult." Based on this feedback, our further research will consider incorporating images along with text.

## References

- Sungeun An, Robert Moore, Eric Young Liu, and Guang-Jie Ren. 2021. [Recipient design for conversational agents: Tailoring agent’s utterance to user’s knowledge](#). In *Proceedings of the 3rd Conference on Conversational User Interfaces, CUI ’21*.
- Richard Desjardins, William Thorn, Andreas Schleicher, Glenda Quintini, Michele Pellizzari, Viktoria Kis, and Ji Eun Chung. 2013. *Oecd skills outlook 2013: First results from the survey of adult skills*.
- Philippe Laban, John Canny, and Marti A. Hearst. 2020. [What’s the latest? a question-driven news chatbot](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 380–387.
- Philippe Laban, Elicia Ye, Srujay Korlakunta, John Canny, and Marti Hearst. 2022. [Newspod: Automatic and interactive news podcasts](#). In *Proceedings of the 27th International Conference on Intelligent User Interfaces, IUI ’22*, page 691–706.
- OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, and Shyamal Anadkat. et al. 2024. [Gpt-4 technical report](#). *Preprint*, arXiv:2303.08774.
- Ian Stewart and Rada Mihalcea. 2022. [How well do you know your audience? toward socially-aware question generation](#). In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 255–269.