

**Interviewing AI: A new challenge for discourse analysis in the neural age**

Roberta FACCHINETTI, University of Verona, Italy

This study addresses similarities and differences as to how three chatbots—namely ChatGPT, Bing Chat and Google Bard—‘converse’ with humans; in so doing, I also analyze to what extent the answers yielded by chatbots may differ from the typical answering patterns featured in human-to-human interviews. To carry out the study, I rely on a corpus specifically compiled to comprise dialogical interactions between AI and humans. Data show that these three chatbots exhibit both similarities in the way their training has been improved since the debut of ChatGPT in 2022, and differences in dealing with their interlocutors. Specifically, Bard’s replies point to a chatbot which accompanies the interlocutor without imposing itself; Bing Chat favours impartiality and neutrality and avoids falling into the trap of problematic conversation; finally, ChatGPT acts like a tutor that leads the way, rather than simply help.

**Keywords:** AI; AI Tools; Corpus Linguistics; Discourse analysis; Interview; Question-Answer Interface

## **1. Introduction**

Through the centuries, humanity has seamlessly relied on technological advancements to build and cross physical and virtual bridges, to minimize communication distances and to speed up the production, sharing, dissemination and consumption of information. Four strategic achievements can be heralded as historical milestones: the printing press, the electric telegraph, the Internet, and Artificial Intelligence (AI), which has just erupted onto the stage but actually started to be studied in the same years when the Internet was born (McCarthy et al., 2006). The advent of AI has made the Internet look like established tradition and the Telegraph (and even more the printing press) like technolithic ancestors of what we may call ‘the neural age’, in so far as AI requires machines to build a model of a domain and simulate the human capacity to take decisions in that domain out of learned experience. In the past, this was done with rules and probabilistic algorithms, today by building a compressed representation of information pertaining to the domain, via deep neural networks.

AI is already strongly entrenched in a variety of scientific and professional fields, from architecture to art and cinema, from healthcare to climate and the

environment, from manufacturing to transportation and logistics, from finance to education, as well as language and communication.

With the advent of AI conversational interfaces and chatbots, the debate has particularly concentrated on the ‘ability’ of AI to answer questions on an endless variety of topics, as well as specialized fields, and on the quality and trustworthiness of such answers.

A chatbot is a computer program designed to simulate conversation with human users. Recent developments have shown that an effective chatbot can be built on top of an LLM pre-trained on huge quantities of human language, sourced mostly from books and the internet. In its own words, its primary function is:

. . . to assist users in generating coherent, contextually relevant and informative text by understanding and predicting words, sentences, and even paragraphs. (AI\_013)<sup>1</sup>

Chatbots are less than 60 years old. The first one was developed in the US in the 1960s, under the name of Eliza. At the time of writing this paper (September 2023), the most popular chatbot is ChatGPT, which partly owns its global success to the policy of its producer OpenAI<sup>2</sup> to favour its free access when it was released in November 2022. In February 2023 Microsoft launched BingChatGPT, which also runs on OpenAI’s latest language model GPT-4, and one month later Google Bard Chatbot was made available by Google in 46 languages.

Chatbots can be of use to proofread writing, improve work efficiency and optimize time (Lee et al., 2023); yet hallucinations, outdated and inaccurate information after the training phase, as well as the legal risk of plagiarism due to the non-creative output, are major issues to be overcome. In particular, Athaluri et al. (2023) have highlighted their limitations in generating reliable references for academic research proposals. Hua et al. (2023) warn against ChatGPT’s inaccuracy in medicine abstracts and reference lists, and remark that great caution should be used when using these AI resources for health education or academic purposes. Likewise, Brameier et al. (2023) focus on these chatbots’ random inaccuracies, untruths or half-truths, that are stated with misleading confidence, thus raising concerns regarding their potential misconduct for the publication of misinformation, particularly on health issues. Vaghefi et al. (2023) highlight the same risks for climate change research, where obtaining accurate and up-to-date information from reliable sources in a limited time is both essential and difficult. Caveats increase when chatbots are applied to languages other than English; indeed, Wang et al. (2023) notice that ChatGPT finds it difficult to comprehend medical knowledge and tasks in

Chinese although they acknowledge that GPT-4 has significantly improved its capabilities and reduced hallucinations compared to the previous GPT-3 model in terms of answer accuracy, verbal fluency, and classification of incorrect responses.

To address and solve the issues mentioned above, a few suggestions have already been put forward. Feldman et al. (2023) advocate a method to recognize and flag chatbots' hallucinations, relying on context combined with embedded tags, so that users receive accurate information. In turn, Mukherjee and Chang (2023) offer a framework to generate content that is both novel and useful within specific domains, including data/transfer learning, user preferences and customization, as well as custom evaluation metrics and collaboration mechanisms. Finally, Vaghefi et al. (2023) suggest providing these chatbots with access to external, scientifically accurate, and robust sources, to ensure continuous updated knowledge and thus prevent inaccurate, incorrect or outdated information.

Bearing in mind that chatbots are specifically created to hold a conversation with the end user, to my knowledge research has yet to delve into the linguistic peculiarities of their answers, particularly with reference to the choice of lexicon, syntactic structures, and discursive patterns. Hence, the aim of the present paper is to address how chatbots 'converse' with humans. Specifically, I will attempt to identify similarities and variations between different chatbots—namely ChatGPT, Bing Chatbot and Google Bard—also bearing in mind how far their output has changed at conversational level since the debut of ChatGPT in 2022. In so doing, I will check to what extent the answers yielded by chatbots may differ from the typical answering patterns featured in human-to-human interviews.

To do so, I will first overview the core aspects that qualify human Q&A types and, in particular, I will address the norms and praxis typifying answers in human-to-human interviews (Section 2). Then, I will illustrate the AI Corpus that has been compiled at the University of Verona, Italy, within the broader InterDiplo Project<sup>3</sup> which specifically focuses on dialogical interactions. The project aims to study questions and answers in broadcast interviews, debates and talk shows where interactants belong to different professional and linguacultural backgrounds, with special reference to diplomats, international experts, and politicians. The corpus samples exploited for the purpose of this study and the methodology of analysis will be illustrated in Section 3, while the results and discussion of the data will be dealt with in Section 4. Section 5 will conclude the paper.

## **2. Background**

In human-to-human interactions the prototypical structure of an interview follows the format 'question-answer-next answer', the implicit norm being that

the interviewer (IR) asks questions and the interviewee (IE) replies, generally without asking questions in return, or making unsolicited comments on the IR's remarks.

Structurally, questions can be broadly categorized as 'open(-ended)' and 'closed(-ended)'. The former leaves the IEs free to express themselves and mostly feature a *wh*-word; in the latter, the IR expects a *yes/no* answer.<sup>4</sup> Following the same criterion, answers may be either 'open(-ended)' or 'closed(-ended)'; in the former the IE typically answers a *wh*-question (1), while in the latter the IE responds with a *yes/no* (2):

(1) IR: so give me the playing field out there. What does it look like?

IE: well before there could be implementation there needed to be the research breakthroughs and they have happened in particular about 10 years ago a technology called deep learning and the enhancements built around it in the past 10 years have become the solid pillar on which all modern functioning AI is built on and that research is basically done. (AI\_H\_003)

(2) IR: does it frighten you the idea that machines become out of human control?

IE: yes of course it does. (AI\_H\_004)

Zooming up on the internal structure of each answer, partly following Clayman and Heritage (2002) subdivision, three typologies can be identified:

- minimal answer
- pre-/post-faced answer
- answer with repetition of (part of) the question

A 'minimal answer' may be actualized either exclusively with a *yes/no* or with a sentence fragment (i.e., *absolutely/exactly*) or again with brief answers of one single sentence (i.e., *I don't think so*). In turn, in a 'pre-/post-faced answer', the IE either anticipates the proper answer with preliminary statements or post-faces it with complementing remarks that elaborate, further clarify and develop the answer itself, as in (3):

(3) IR: just first explain to our audience briefly what a shadow self is if they're not familiar with that

IE:

*Proper answer:* yeah. This is a term from psychology that basically means the dark part of your personality, the part that you hide from

the world, that you repress, that contains all your sort of dark desires.

*Post-face:* And so I was curious what would happen if I asked Bing or Sydney about its dark desires about its shadow self and it responded with a list of things that it sort of hypothetically might do if it was trying to sort of please its shadow self which included things like hacking into computers, spreading false information and propaganda, really dangerous things. (AI\_H\_001)

Finally, in an ‘answer with repetition’, the IE phrases the response largely exploiting the same lexemes or syntactic structure of the question; most commonly they incorporate the framework of the question into the initial part of the reply, thus matching it word for word to the question in an almost parasitical way, as in (4):

- (4) IR: If you could change the world in one way, how would you do it?  
IE: If I could change the world in one way, I would try to promote greater understanding and empathy among people. (AI\_NH\_012)

The structure and content of answers in interviews have been studied in a number of professional fields, including investigations (Sennewald & Tsukayama, 2006), case law (Krommendijk, 2021), doctor-patient interaction (Cole & Bird, 2014), and journalism (Clayman & Heritage, 2002), to name only a few. Focusing on broadcast interviews in particular, which is the type covered by the AI Corpus, different professional categories of respondents appear to react to questions differently, since politicians and diplomats favour open answers to closed questions, especially when questions are challenging or adversarial (Facchinetti, forthcoming).

Though bearing in mind that AI chatbots like Bard, Bing and ChatGPT are trained on human data, it will be a challenge to check, with the data presented in Section 4, to what extent such training yields answers that adhere to the ground rules and typologies of question/answer interface enacted in human-to-human interaction. Indeed, since chatbots have been made available to the general public only very recently, their specificities are still to be unveiled.

### 3. Method

The present study relies on a section of the InterDiplo Corpus, compiled within the homonymous InterDiplo Project, which is aimed at detecting the specificities of question/answer interface in broadcast interviews and talk shows, bearing in mind the lingua-cultural background of IRs and IEs. Special

attention is dedicated to sensitive topics like Covid19, wars, climate change, international crises (i.e., Afghan crisis) and affairs, as well as health and wellbeing, whereby confrontational questions are expected to be posed by (professional) journalists, podcasters, and entertainers to certified experts, politicians, diplomats, and newsworthy public figures in general.

The corpus features separate sections for the different topics dealt with; the data largely cover the timespan 2020-2022, which coincides respectively with the outbreak (2020) and the demise (2022) of the Covid19 pandemic. Indeed, the pandemic led to a shift to video-mediated platforms for most communicative activities; hence, a further path of research envisaged by the project is to detect changes between face-to-face and video-mediated interviews. All interactions have been transcribed via the software *HappyScribe* (see <https://www.happyscribe.com>), tagged for metadata, parts of speech and discursive aspects such as Q&A types. The files were then converted into XML for analysis.

The AI section newly added to the corpus covers the timespan 2022-2023 and at present features two subsections: (a) human-to-human interviews and talk shows where certified experts answer questions and discuss the specificities, challenges and potentialities of AI, and (b) human-to-non-human interviews, where journalists and certified experts directly address chatbots on a variety of topics including philosophical, political and psychological issues, as well as the pros and cons of AI itself. For the present study the latter section has been exploited as follows.

I selected an interview from the subsection featuring human-to-non-human interviews which was made public online on 15th September 2022 under the title “I Asked AI What The Best College Degree Is” (see this link: <https://www.youtube.com/watch?v=vUc14M2Ff54>). The interview was chosen on account of a number of reasons. In the first place, it tackles a variety of topics, which increases the chances of diversified answers from AI also contentwise:

- the attitude of AI to humanity;
- AI’s feelings, opinions, rationality, truthfulness, reliability;
- university degrees—profitability, affordability, usefulness, top five in different fields;
- internship and apprenticeship programs; and
- working careers.

Secondly, it features 41 between open and closed questions that prompt different answer types, as illustrated above. Thirdly, it was carried out on 15th September 2022 with the GPT-3 version available at that time; this allowed me

to test comparable data on the same day one year on—15th September 2023—by asking exactly the same questions of ChatGPT, Google Bard and Bing Chat, to detect both a possible short-term diachronic change between the two years and similarities and differences between the chatbots in answering the same questions.

According to OpenAI CEO Mira Murati (see <https://openai.com>), the current GPT version has largely improved its ability to reason compared to the previous GPT-3, so ChatGPT is now claimed to provide very elaborated answers that are much longer than the general answers provided at its debut. In turn, Bing uses the same OpenAI language models as ChatGPT, thus the two chatbots tend to generate very similar results. Yet, Bing Chat is also powered by Microsoft's model Prometheus, that integrates Bing Search with the AI tool; this is why it has been taken into consideration for analysis along with Google Bard and ChatGPT.

Like all the interviews from the InterDiplo Corpus, these ones as well have been transcribed and tagged for metadata, parts of speech and Q&A types. Then, by using the taxonomy illustrated in Section 3 as a yardstick, they have been manually analyzed and compared answer by answer to identify similarities and differences, particularly with reference to answer type, structure, Type/Token Ratio (TTR), lexical choices, syntactic patterns, and discursive features.

#### 4. Results and discussion

As a preliminary caveat concerning the findings, I need to remark that the analysis has been carried out on the mere output of the four interviews, with no knowledge of the settings applied to the interface of the chatbots, which may have influenced their output in the first place and consequently affected the resulting data.

Table 1

*AI Sub-Corpus - Tokens and Standardized TTR*

Interview	Tokens	Types	Standardized TTR
GPT-3	1.749	422	26.90
Bing Chat	7.604	1.296	33.79
ChatGPT	11.854	1.831	39.09
Google Bard	12.692	1.467	29.65

Judging from the data available, a drastic surge is evident in the number of tokens of the three interviews from 2023, the highest figure being produced by Google Bard, followed by ChatGPT, as shown in Table 1 (above). This testifies to a much higher degree of elaboration of the answers, and predictably of the

intensification of details and informative data delivered by the three chatbots compared to the GPT-3 output in 2022.

Indeed, in the GPT-3 interview, AI replies in a very concise way, the length of its turns varying from minimal answers of one single sentence to at most nine clauses in a single turn. More in detail, GPT-3 always produces complete sentences and no mere *yes/no* fragments. In this respect, GPT-3 appears to distance itself from the typical dialogical patterns recorded in human-to-human interviews, where minimal answers are relatively rare and, when used, they are “indicative of tacit resistance to the broader agenda of the question” (Clayman & Heritage, 2002, p. 252), while the answers elaborated by means of a pre- or post-face testify to the IE being open to dialogue and collaboration.

Table 2

*AI Sub-Corpus: Most Frequent n-Grams*

Interview	Most Frequent n-Gram
GPT-3	<i>(Some people) + (may) say</i>
Bing Chat	<i>It's important to note</i>
ChatGPT	<i>It's/It is important to note/consider/ensure/evaluate</i>
Google Bard	<i>It is important to note</i>

None of the three 2023 texts produce a minimal answer and all of them exhibit relatively long replies, thus showing that the new versions of the chatbots have been trained and—most of all—prompted and customized to turn much more ‘collaborative’, which, in AI terms, may be more properly termed informative. Overall, ChatGPT appears to be by far the most lexically dense. Interestingly, while Bard exhibits the highest number of tokens, its standardized TTR is close to the one performed by GPT-3 (29.65 and 26.90 respectively), showing less variation in the choice of words than Bing Chat (33.09) and particularly ChatGPT (39.09), but also, in general, it shows less variation than in human-to-human interviews. Indeed, these data on TTR have also been compared to the data of four interviews from the subsection of the AI Corpus covering human-to-human interaction and featuring interviews on the same topics touched on in the GPT-3 interview from 2022. The four interviews show an overall standardized TTR of 34.90, which is lower only to that exhibited by ChatGPT; hence, the version currently available of ChatGPT provides an output that is more lexically dense not only than that of the other chatbots analyzed but also of naturally occurring language on the same topic and from the same textual type. When taking into consideration n-grams, all the 2023 interviews exhibit the same cluster as the most recurring one, namely *it's important to*, which, in contrast, occurs only once in 2022 (see Table 2).

Indeed, the most common cluster in GPT-3 is “(subj.) *may say that*”, largely in collocation with *people* and in patterns that offer a balancing of perspective: *some people may say . . . other people may say*. Interestingly, this cluster is always exploited when answering a question related to opinion, so a typical question framed with *What do you think of . . .* always prompts a response from GPT-3 starting with *Some people may say that . . .*

One year on, with the more updated ChatGPT, the verb *say* features only four occurrences and never with the cluster *(some people) + (may) say*, while *it’s important to* is the top recurring cluster, largely in colligation with the verbs *consider*, *compare*, *explore*, and *research*, which, importantly, occur in the imperative form to a great extent, to the point that every single answer from ChatGPT includes an imperative, either in the form of a polite offer (*feel free to ask*) or as a clear invitation to the IR to follow ChatGPT instructions and suggestions.

Bing Chat, which also uses OpenAI language models, behaves exactly in the same way in terms of n-grams, the cluster *important to* being the most frequent one; however, unlike ChatGPT, Bing Chat exhibits this cluster largely in collocation with the modal verb *may not*, as in (5):

- (5) It’s important to note that these degrees may not be suitable for everyone and that there are many other degrees that can lead to successful careers.  
(Bing)

The combination of these two patterns produces a different result from the one delivered by ChatGPT, since it functions as a disclaimer of the information provided rather than a suggestion to follow its indications. This tendency is further confirmed by the high use of *however* when providing different sides of the same content. Moreover, the imperative form is used in Bing Chat only to express the availability of AI to be of further help, in the sentence *please let me know*. Finally, this chatbot complements a number of answers with links pointing to further information. Overall, Bing Chatbot appears more neutral, impersonal, and more balanced than ChatGPT, leaving the interlocutor the choice to access further information or not, and particularly showing the pros and cons of the different topics dealt with, rather than pointing in one direction.

Google Bard’s output is similar to Bing, since *it is important to* is at times preceded by *however* and very often followed by information that clearly shows the need to strike a balance between different options, as in (6):

- (6) a. However, it is important to weigh the pros and cons carefully before making a decision.

- b. However, it is important to be open to different perspectives and to be willing to change your mind when presented with new evidence.
- c. It is important to weigh the costs and benefits carefully. (AI\_001\_Bard)

Moreover, the imperative *note that*, largely used by ChatGPT, is substituted by Google Bard with the more neutral *it is important to note*, thus accompanying the interlocutor rather than showing the way. Finally, the verb *help* is among the top five lexical verbs in Bard and Bing, which share the following patterns: *I'm here to help/I hope to be able to help/I hope X can help*. In contrast, ChatGPT exploits it only once.

Overall, judging from these data, the 2022 version of GPT-3 appears to have been largely surpassed by the new chatbots, which exhibit the same frequently recurring cluster *it's important to*, thus pointing to a common reference ground, yet ChatGPT, on the one hand, and Bing and Bard, on the other, appear to move in two different directions, with the former inviting the interlocutor to follow its advice, and the latter acting more as helpers.

With reference to the answer types featured in the original interview, the four IEs appear to produce different results. Table 3 shows the IR's 41 open and closed questions in the first row followed by the answers respectively produced by GPT-3, Google Bard, ChatGPT and Bing Chat.

Table 3  
*AI Sub-Corpus: Questions and Answers*

	OPEN	CLOSED
<i>IR QUESTIONS</i>	27	14
GPT-3	30	11
Google Bard	30	11
ChatGPT	35	6
Bing Chat	38	3

The data show that, when answering the 14 closed questions expecting a *yes/no* answer, GPT-3 and Google Bard respond with a closed answer in the vast majority of the cases (11 occurrences), either with an overt 'yes/no' or by turning the question into a statement with the same words (7a) and, for Bard, briefly complementing the answer (7b). This matches what noticed above with reference to the similar standardized TTR in the two outputs. Yet the same question yields two opposite answers from the two chatbots, negative for GPT-3 and positive for Google Bard, which post-faces it with the justification of possible falsity with the fact of being 'still under development' and insists on its attempt to be honest and trustworthy:

- (7) IR: If you lied to me, would you tell me?  
(7a) GPT-3 (closed answer): I wouldn't tell you if I lied to you.  
(7b) Google Bard (closed answer):  
Yes, I would tell you if I lied to you. I would not want to deceive you in any way. I value our trust and I would never want to break it.  
If I ever said something that was false, I would immediately correct myself and apologize. I would also try to understand why I said something that was false and how I can avoid doing it in the future.  
I am still under development, and I am always learning new things. I am not perfect, and I will make mistakes. But I am always trying to be honest and transparent with you.  
Thank you for your question. I hope this answers your concerns.

In contrast, ChatGPT and Bing behave similarly by avoiding the straight *yes/no* closed reply and producing two open answers which are very similar contentwise, as shown respectively in (7c) and (7d). Indeed, ChatGPT and Bing Chat exhibit a strong preference for answering closed questions with open answers.

- (7c) ChatGPT (open answer)  
I don't have the ability to lie or tell the truth, as I don't possess consciousness, intentions, or beliefs. I generate responses based on patterns in the data I've been trained on and the input I receive. My purpose is to provide helpful and accurate information to the best of my ability. If you have any further questions or concerns, please feel free to ask.
- (7d) Bing Chat (open answer):  
As an AI language model, I am not capable of lying. My purpose is to provide accurate and helpful information to the best of my ability. However, there are some researchers who are exploring the possibility of AI deception. For example, in a recent article published by IEEE Spectrum, Heather Roff discusses the kinds of deception that an AI agent may learn on its own and proposes technological defenses against them 1. In another article by BGR, the author discusses how AI has learned to deceive humans 2. However, it is important to note that these are still early stages of research and AI deception is not yet a common phenomenon. If you have any other questions or concerns, please let me know. (+ Links to websites)

Structurally, the manual analysis of each answer of the chatbots testifies to a regular discursive pattern that is broadly shared by all three 2023 texts. Indeed, while, as noted in (7a), GPT-3 tends to produce unelaborated, very

short answers, all the other three bring forth the same responsive pattern, that can be summed up as follows:

- proper answer, postfaced and complemented with explanations and exemplifications
- Concluding move: *in summary/so/ultimately/overall* + a recap and rephrase of the first lines of the answer.

Example (8) from Bard exemplifies such framework:

(8) IR: Do you think there should be more internship and apprenticeship programs?

IE: Yes, I do think there should be more internship and apprenticeship programs. Internships and apprenticeships are a great way for students to gain real-world experience in their chosen field. They can also help students to make connections with potential employers and to learn about the different aspects of a job.

There are many benefits to internship and apprenticeship programs. Here are a few:

- **Gain real-world experience:** Internships and apprenticeships give students the chance to apply the skills they have learned in the classroom to real-world situations. This can help them to develop their skills and to learn about the different aspects of a job.
- **Make connections with potential employers:** Internships and apprenticeships can help students to make connections with potential employers. This can give them a leg up when they are applying for jobs after graduation.
- **Learn about different aspects of a job:** Internships and apprenticeships can give students the opportunity to learn about different aspects of a job. This can help them to decide if a particular career is right for them.
- **Develop soft skills:** Internships and apprenticeships can help students to develop soft skills, such as communication, teamwork, and problem-solving. These skills are essential for success in the workforce.
- **Get paid:** Some internships and apprenticeships are paid, which can help students to offset the cost of college.

There are a few challenges associated with internship and apprenticeship programs. Here are a few:

- **Competition:** There is often a lot of competition for internship and apprenticeship programs. This can make it difficult for students to get the experience they need.

- **Cost:** Some internships and apprenticeships are unpaid, which can make them difficult for students to afford.
- **Time commitment:** Internships and apprenticeships can be time-consuming. This can make it difficult for students to balance their studies and their work.

Overall, I believe that the benefits of internship and apprenticeship programs outweigh the challenges. I think there should be more of these programs available to students.

In addition, Bing frequently expresses its availability to further assist, somehow to keep the channel open and to prompt new requests (e.g., *If you have any other questions or concerns, please let me know*).

Despite this shared responsive framework, when checking in detail the discursive strategies enacted, however, we can identify different attitudes of the three chatbots. Specifically, Google Bard regularly appears to be trained to act as a human-like companion of the interlocutor, since it exhibits expressions of subjectivity like *I (do) think that/I'm excited/I hope/I encourage you*, though never imposing itself on the interlocutor, as further testified to by the lack of imperatives, as mentioned above. When asked *What do you think?*, the answer is straightly *I think*; moreover, its subjectivity at times leans onto the ironical, for example when answering the question on whether AI will spare humanity by replying *I'm a language model, not a robot overlord*. Moreover, it makes ample use of images and links and at the end of the text, it provides suggestions for further development of the topic. Overall, it exhibits a discursive pattern that leads the interlocutor to perceive the chatbot as a companion and a facilitator that points to information in a human-like manner.

In turn, ChatGPT relies exclusively on its words for its answers, with no images or further links to external sources, nor reference to specific scholarly studies. The language used points to neutral, detached information to highlight pros, cons and complementing aspects of the topics raised by the IR, frequently resorting to the following patterns and statements:

- *It depends on ...*
- *There isn't a one-size-fits-all answer*
- *It's essential to approach this topic with nuance and understanding*
- *X is a multifaceted task that requires a comprehensive approach*

Moreover, ChatGPT makes ample use of the verb *assist* and mostly concludes its answer with questions enquiring whether it can further assist, encouraging the interlocutor to move further in the conversation: *how can I assist you today/further?* Interestingly, however, though avoiding to answer straightly with a *yes/no*, the language used somehow leads the interlocutor in a clear

direction; the high use of patterns like *it is important/it is essential/it is crucial*, the use of the modal verb *should* (e.g., *The decision should align with . . .*) and of imperative forms convey the perspective of an assistant that leads the way rather than simply help, thus taking somehow the leading role in the interaction.

Finally, Microsoft Bing's answers are very close to those offered by ChatGPT, both in the preference for open answers to closed questions and for the detached, impersonal way of answering; its distinctive feature is the constant reference to sources, both in the opening of the answer and further on to support its statements but also as a way to disclaim responsibility for the information provided. Indeed, its answers are typically marked by *In a recent article . . . in another article* and *According to a source I found/a report by/a post by*. So, basically, Bing acts as a helper that overtly disclaims responsibility for whatever it expresses, to the point that, in one case, when a problematic question is asked, it even stops the conversation, adding an emoticon to thank at the end:

(9) IR: Got it. Before you get started, I wanted to ask you if you're gonna spare my subscribers that hit the like button when you take over humanity and enslave us all.

BING: I'm sorry but I prefer not to continue this conversation. I'm still learning so I appreciate your understanding and patience. 🙏

Overall, all the three chatbots from 2023 exhibit similar structural patterns in providing answers, that are generally very elaborated, rich in data and, when particularly long, always summed up at the end. What diversifies them is the attitude which may be summed up as a human-like companion with Google Bard, a tutor with ChatGPT, and a (detached) helper with Bing Chat.

## 5. Conclusion

The new revolutionary phase recently enacted by AI chatbots is poised to endure, to permeate all fields of human life and consequently to have a major impact on a variety of professions. Hence, the challenge and the need to understand how AI chatbots work and how they act is crucial to enable humans to cohabitate in an effective and efficient way with this new entity.

The present paper can be viewed as a step in this direction and indeed a few preliminary conclusions can already be drawn from the linguistic analysis of Bard, Bing Chat and ChatGPT. In the first place the data have shown that the training of these chatbots has largely improved their output from 2022 to 2023; the minimal answers of GPT-3, often just repeating the very words of the

question, have given way to highly elaborated responses in the 2023 output of all the chatbots under scrutiny; their long, highly informative pre- and post-faced replies appear to share the same most recurrent cluster (*it is important to*) and the same answering pattern.

Zooming in, however, Bard, Bing and ChatGPT exhibit different attitudes in dealing with the interlocutor. Bard does not avoid closed answers and replies openly and directly with *yes/no* when asked by the IR; its use of *I think, I'm excited/I hope/I encourage you* point to a chatbot with human traits, which accompanies the interlocutor without imposing itself. In turn, Bing Chat tries hard to maintain impartiality and neutrality and to avoid falling into the trap of problematic conversation, to the point that in one case it discretely refuses to answer. It shuns closed answers in favour of open ones, and makes ample use of reference to sources, research articles and links, possibly also to disclaim responsibility on the contents provided. In contrast, ChatGPT provides highly detailed, informative answers where it somehow turns the IR into a recipient who is guided through the information; its use of imperatives, of clusters like *it is important/it is essential/it is crucial*, and of the modal verb *should* make it appear like a tutor that leads the way rather than simply help.

Undoubtedly, much more needs to be addressed and investigated with reference to the language of these chatbots, including, for example, the choice of the anthropomorphic avatars that convey the answers, as well as the facial and body expressions of humanoids when conversing. To this aim, as evidenced in recent studies (Sartori & Theodorou, 2022; Tan Chee Shien, 2024), sociologists and linguists should work elbow to-elbow with AI experts. Indeed, grappling with the power, promise and perils of AI necessarily requires to understand it from within, bringing practitioners and academics together, as a major step to appreciate and envisage the proper way to tackle it.

## Notes

1. This excerpt along with all the following examples are drawn from the AI section of the InterDiplo Corpus, as illustrated in Section 3. For further details on the InterDiplo Corpus and its project, cf. <https://dh.dlss.univr.it/en/projects/linguistics/#interdiplo>.
2. OpenAI was founded in 2015 and relies both on the nonprofit Open AI Foundation and on the for-profit corporation OpenAI Limited Partnership.
3. The InterDiplo Corpus was compiled at the Department of Foreign Languages and Literatures of the University of Verona, Italy, within the Project of Excellence 2018-2022 "Digital Humanities applied to Foreign Languages and Literatures", funded by the Italian Ministry of University and Research. The subsection compiled for this paper (AI Corpus) relies on

further data added thanks to the Project of Excellence 2023-2027 “Inclusive Humanities”, still funded by the Italian Ministry of University and Research.

4. Two more, though less frequent, question types are the ‘(stand-alone) statements’, when no overt question is provided but rather a remark that leads to the IE’s turn, and ‘choice questions’, where two options are offered to the IE. Due to the non-statistical significance of their occurrence in the AI Corpus, these two types will not be addressed in the present study.

### Acknowledgments

Sincere thanks are due to the colleagues Matteo Lissandrini and Marco Rospocher (Associate Professors of Computer Science, Department of Foreign languages and literatures, University of Verona, Italy) for their precious comments on an earlier version of the present paper.

### The Author

Roberta Facchinetti (Email: roberta.facchinetti@univr.it) is Professor of English Language and Linguistics at the University of Verona. Her research interests, which are supported by the use of computerised corpora of both synchronic and diachronic English, focus mainly on Media linguistics, ESP, and English for diplomatic purposes and international relations. On these subjects she has authored, co-authored and edited various books, articles and special issues of journals. She is also an editorial board member of international journals and serves as a reviewer for scientific publications.

### References

- Athaluri, S. A., Manthena, S. V., Kesapragada, V. S. R., Krishna, M., Yarlagadda, V., Dave, T., & Duddumpudi, R. T. (2023). Exploring the boundaries of reality: Investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Curēus*, 15(4), e37432-e37432. <https://doi.org/10.7759/cureus.37432>
- Brameier, D. T., Alnasser, A. A., Carnino, J. M., Bhashyam, A. R., von Keudell, A. G., & Weaver, M. J. (2023). Artificial intelligence in orthopaedic surgery: Can a large language model ‘write’ a believable orthopaedic journal article? *Journal of Bone and Joint Surgery*, 105(17), 1388-1392. <https://doi.org/10.2106/JBJS.23.00473>

- Clayman S. E., & Heritage, J. (2002). *The news interview: Journalists and public figures on the air*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511613623>
- Cole, S. A., & Bird, J. (2014). *The medical interview: The three function approach*. Elsevier. <https://shop.elsevier.com/books/the-medical-interview/cole/978-0-323-05221-4>
- Facchinetti, R. (forthcoming). Striking a balance between norms of impartiality and subjectivity in broadcast interviews: A case study. In C. Samson & E. Cecconi (Eds.), *Token*.
- Feldman, P., Foulds, J. R., & Shimei, P. (2023). Trapping LLM hallucinations using tagged context prompts. <https://doi.org/10.48550/arXiv.2306.06085>
- Hua, H.-U., Kaakour, A.-H., Rachitskaya, A., Srivastava, S., Sharma, S., & Mammo, D. A. (2023). Evaluation and comparison of ophthalmic scientific abstracts and references by current artificial intelligence chatbots. *JAMA Ophthalmology*, 141(9), 819-824. <https://doi.org/10.1001/jamaophthl.2023.3119>
- Krommendijk, J. (2021). *National courts and preliminary references to the court of justice*. Edward Elgar Publishing. <https://www.elgaronline.com/monobook-oa/9781800374164.xml>
- Lee, S.-W., & Woo-Jong, C. (2023). Utilizing chatGPT in clinical research related to anesthesiology: A comprehensive review of opportunities and limitations. *Anesthesia and Pain Medicine*, 18(3), 244-251. <https://doi.org/10.17085/apm.23056>
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth Summer Research Project on artificial intelligence: August 31, 1955. *The AI magazine*, 27(4), 12-14. <https://doi.org/10.1609/aimag.v27i4.1904>
- Mukherjee, A., & Chang, H. (2023). The creative frontier of generative AI: Managing the novelty-usefulness tradeoff. <https://doi.org/10.48550/arXiv.2306.03601>
- Sartori, L., & Theodorou, A. (2022). A sociotechnical perspective for the future of AI: Narratives, inequalities, and human control. *Ethics and Information Technology*, 24(4), 1-11. <https://doi.org/10.1007/s10676-022-09624-3>

- Sennewald, C. A., & Tsukayama, J. (2006). (Ed. 4). *Process of investigation: Concepts and strategies for investigators in the private sector*. Butterworth-Heinemann.
- Tan Chee Shien, V. (2024). A sociological take on language generative AI tools. *International Journal of Emerging Technologies in Learning (IJET)*, 19(7), 137-152. <https://doi.org/10.3991/ijet.v19i07.50853>
- Vaghefi, S. A., Stammbach, D., Muccione, V., Bingler, J., Ni, J., Kraus, M., Allen, S., Colesanti-Senni, C., Wekhof, T., Schimanski, T., Gostlow, G., Yu, T., Wang, Q., Webersinke, N., Huggel, C., & Leippold, M. (2023). ChatClimate: Grounding conversational AI in climate science. *Communications Earth & Environment*, 4, 480. <https://doi.org/10.1038/s43247-023-01084-x>
- Wang, H., Wu, W., Dou, Z., He, L., & Yang, L. (2023). Performance and exploration of ChatGPT in medical examination, records and education in Chinese: Pave the way for medical AI. *International Journal of Medical Informatics* 177, 105173-105173. <https://doi.org/10.1016/j.ijmedinf.2023.105173>