

Jie Han · Wei Qiu · Eric Lichtfouse

# ChatGPT in Scientific Research and Writing

A Beginner's Guide



Springer

# Contents

<b>ChatGPT in Scientific Research and Writing: A Beginner's Guide..</b>	<b>1</b>
1 Introduction .....	1
2 Extracting Key Points or Specific Information from Research Papers .....	2
3 Interpreting Figures and Correlating to Specific Conclusions.....	9
3.1 Hyperspectral Images with Spectra.....	10
3.2 Data Plots .....	10
4 Evaluation of Research Papers.....	16
4.1 Reviews versus Original Research Articles.....	17
4.2 Current Policies .....	21
5 Spotting Errors .....	25
5.1 Scientific Errors.....	25
5.2 Misquotation .....	28
6 Responding to Reviewer's Comments.....	30
6.1 Rebuttal to Criticism.....	30
6.2 Answering Questions .....	35
6.3 Responding to Difficult Requests .....	39
6.4 A Reminder for Journal Editors and Reviewers.....	42
7 Advanced Language Editing.....	42
7.1 Correcting Issues in Writing.....	42
7.2 Rewriting Entire Text.....	43
7.3 Removing Language Barriers .....	46
7.4 GPT-3.5 versus GPT-4 Model.....	48
8 Crafting Article Titles .....	48
9 Design of Experimental Studies.....	50
9.1 Designing Experiments and Refining Methods.....	50
9.2 Reality Check .....	53
9.3 Cautionary Note .....	56

10	Design of Public Surveys .....	59
10.1	Creating a Survey Questionnaire from Scratch.....	59
10.2	Refining Scopes .....	63
10.3	Designing a Survey-Based Risk Assessment Study.....	67
11	Writing Research Proposals .....	72
11.1	Brainstorming .....	72
11.2	Writing a Mock Proposal .....	73
11.3	Compliance Requirements .....	83
12	Science Communication and Public Engagement.....	84
12.1	Adapting Research Papers into Various Styles of Writing....	84
12.2	Magazine-Style News Articles and Social Media Posts .....	84
12.3	Creating Visuals .....	85
13	Pitfalls .....	93
14	What Scientists and Developers Could Do.....	95
14.1	Scientists as Early Adopters of AI Large Language Models .	95
14.2	Functionalities in Need.....	101
14.3	Scientific Data for Training Models.....	102
14.4	Prompt Example Sets.....	102
15	Conclusion .....	103
	References .....	103
	<b>Appendix</b> .....	<b>111</b>

# ChatGPT in Scientific Research and Writing: A Beginner's Guide



**Abstract** The developers of ChatGPT have predicted that, within the next ten years, artificial intelligence (AI) systems will exceed expert skill levels in most domains, and carry out as much productive work as one of today's largest corporations. Since the public release of ChatGPT, there has been surging interest in exploring the use of large language models, including ChatGPT, in scientific research, publication, and science communication in general. In this book, we will explore the models' capabilities, including GPT-4, GPT-3.5, and GPT-enabled new Bing (now Copilot), for carrying out the tasks through different stages of scientific research from research conceptualization, study design, to publication and science communication. We used these models for abstracting key points and extracting specific information from research publications, interpreting figures, evaluating research papers, spotting errors, responding to reviewer's comments, language editing, designing experiments, creating survey questionnaires, brainstorming, writing research proposals, and creating visuals. Major limitations of these models include hallucinations, randomness in answers when prompted by identical questions, and the lack of support for big data scrapping, processing, and visualization.

## 1 Introduction

The advancement of artificial intelligence technologies provides scientists with increasingly powerful and accurate research tools. On 14 March 2023, OpenAI released the GPT-4 model, the successor to ChatGPT based on GPT-3.5, which generated extensive discussions in the scientific community (Bockting et al. 2023; Owens 2023; Stokel-Walker 2023; Stokel-Walker and Van Noorden 2023). Five months later, Microsoft announced that over one billion chats and 750 million images had been generated by users within six months since their release of the new Bing (now Copilot), an artificial intelligence (AI)-enabled Internet search engine (Microsoft 2023a). While some argued that ChatGPT's benefits for scientific research are limited, academic publishers and journal editors have responded to the growing proliferation of generative artificial intelligence (AI) models in scientific research



**Fig. 1** Large language models and other artificial intelligence applications, including generative pre-trained transformers (GPTs), show tremendous potential in helping researchers tackle their daily tasks and challenges, such as literature analysis and scientific writing. The image above has been generated by the Bing Image Creator, a deep learning model (DALL-E) generating digital images by natural language descriptions, or “prompts”. Note the slight distortion on the human hand, a characteristic defect that artificial intelligence (AI) models fall short of when creating artistic drawings compared to human artists.

by releasing interim guidelines on their appropriate use in publications (Anonymous 2023a; Elsevier 2024a, b; Thorp 2023; Thorp and Vinson 2023). Here we review the use of ChatGPT (GPT-3.5 and GPT-4) and new Bing in scientific research and science communication. The following sections explore the use of these models for extracting relevant information from research papers, interpreting figures, evaluating research papers, spotting errors, responding to reviewer’s comments, advanced language editing, removing language barriers, adapting manuscripts to various styles of writing, experimental design, creating survey questionnaires, crafting titles, brainstorming, writing research proposals, and creating visuals (Fig. 1).

## 2 Extracting Key Points or Specific Information from Research Papers

Scientists are constantly “under pressure” (Schnoor 2014). As the number of publications grows in virtually every scientific research discipline, there is a pressing need more than ever for researchers to read fast and more efficiently. There are university library guides on speedreading techniques (Cambridge University Libraries 2024;

UoTC 2024). During the COVID-19 pandemic, public health organizations like the World Health Organization found it challenging to keep up with the new research and evidence that emerged every day (Feuer et al. 2020). But just how many papers could one possibly read within a certain period of time? It is a hard question to answer even for the most dedicated researchers.

Using ChatGPT, one could effortlessly delve into the full text of research papers to extract the key findings or specific information queried by the user. One can achieve this with consistently good results using concise instructions, i.e., without crafting multiple prompts with stepwise instructions. In the first example below, we asked the new Bing, a ChatGPT-enabled version of the Microsoft Bing search engine, to analyze the full-text PDF of a short article written by one of the authors. The prompt is shown at the top of the conversation in Table 1. The “More Creative” mode, which generates “Responses (that) are original and imaginative, creating surprise and entertainment for you”, was used in this example.

It is worthwhile to note that the short article was written in a didactic style with discussions throughout the entire text. The article does *not* have a regular structure containing the “Abstract”, “Conclusion”, “Highlights”, or any summarizing text, thereby giving no hint to the AI large language model on *what* to look for in the full text. The meticulous requirements specified by the user in the prompt, i.e., methods, implications, novelty, and significance, pose further challenges for the model to extract such information from the full text because there are no subsections addressing

**Table 1** The key points of a research paper by Han and He (2021a) using the ChatGPT-enabled new Bing under the “more creative” mode

<p><b>Prompt:</b> Summarize the key points of this article in bullet points. Tell me the key findings, methods (how authors obtained their results), implications, and the novelty and significance of this study.</p>
<p><b>Response:</b> Here are the key points of the article in bullet points:</p> <ul style="list-style-type: none"><li>• The article discusses the potential health risks of inhaling micro(nano)plastic debris from masks, respirators, and home-made face coverings during the COVID-19 pandemic.</li><li>• The authors examined several top-selling medical face masks and N95 respirators in China using optical and scanned electron microscopy and found abundant loose debris on their inner facings, some of which could be inhaled by the wearer.</li><li>• The authors also pointed out the possible sources of respirable debris from home-made cloth face coverings, such as fabrics, detergent residues, lint, and household items used as filters.</li><li>• The authors called for collaborative efforts from scientists, manufacturers, and regulators to assess the risks and reduce the micro(nano)plastic and other respirable debris in face masks and respirators worn by millions of people worldwide during the pandemic.</li><li>• The article is novel and significant because it raises an important issue that has been neglected in the current standards and research on respiratory protection devices. It also provides some preliminary evidence and suggestions for future studies on this topic.</li></ul>

Note that the article does not have an “Abstract” or a “Conclusion” section, making it impossible for the model to “cheat its way out” by paraphrasing text summaries that are typically present in these sections of a research paper.

these topics in the article. After validating the model's response, we concluded that the model had delved into the full text with a good understanding of the article and accurately abstracted the key findings, as one would generally agree with after reading this article.

In the first example, the ChatGPT-powered new Bing showed human-level intelligence by having a good understanding of the user's intent and analyzing the full text of a short research paper written in a didactic style. The model provided answers to the user in an accurate, coherent, and context-relevant manner. These initial results were encouraging because completing such tasks would usually require human intelligence, e.g., high school to college student level. Since it was a short research paper addressing a common issue in the public health domain, we wondered how the AI model would perform on lengthy research papers in more specialized research domains.

We then repeated our test on a regular-length research paper using the new Bing under the same setting (Table 2). In this example, we focused on testing the model's capabilities in abstracting key findings and finding out how authors obtained these specific findings. These are usually the first step for one to get an understanding of the essentials of a research paper, before getting into more details. The paper, which has about 6100 textual words, four figures, and one data table, dealt with an unreported compound-polymer duo with potential implications in environmental chemistry research disciplines (Han et al. 2015).

In this example, the model went beyond the Abstract and Conclusion as well as the captions of figures and tables by analyzing the full text of this paper. Again, the model provided accurate, context-relevant answers that are on par with human intelligence. This extraordinary capability of the large language model allows users to quickly extract the key findings and methods from lengthy research papers in less than a minute, compared with the traditional routine of human analysis, i.e., by reading the title → abstract → conclusion → figure(s) and table(s) → experimental, a technique we teach our students in class for speedreading research papers. Note that in this example, we used one natural-language command, i.e., *“List the key findings in this article in detail with bullet points, with information on how authors obtained each of these findings”*, without providing step-by-step instructions to break down the tasks for the model. The model adapted to this request by giving answers in bullet points structured into two different levels. The two-level bullet point layout is clearer and more amenable to fast reading, e.g., when compared with reading the texts in the Abstract and Conclusion.

These results could offer tremendous benefits for scientific researchers, for whom reading research papers is part of the routine tasks in their daily work. We wanted to find out if large language models could consistently provide good results in this domain, e.g., for environmental scientists. One of the best ways of testing the model's capability of abstracting key findings and methods from research papers is to have the model analyze one's own work. Since authors know their papers inside out, that is, both the writing and the research behind it, they have good judgment on the model's true performance in extracting the most valuable information from their papers.

**Table 2** The key findings of a regular-length research paper summarized by the ChatGPT-enabled new Bing under the “Creative Mode”, with insights into how authors obtained each of the key findings

---

**Prompt:** List the key findings in this article in detail with bullet points, with information on how authors obtained each of these findings.

---

**Response:** Here are some key findings in this article in detail with bullet points:

- The article reports the unusual sorption capabilities of polyurethane foam (PUF) for bisphenol A (BPA), a major industrial plasticizer and endocrine-disrupting chemical.
    - The authors obtained this finding by conducting batch sorption experiments and comparing PUF with two benchmark adsorbents, activated carbon (AC) and macroreticular adsorbent resin (dvb-PS).
  - The article shows that PUF has not only anomalously high sorption capacities, but also rapid uptake for BPA compared with AC and dvb-PS.
    - The authors obtained this finding by fitting the sorption data to appropriate isotherm and kinetic models, and measuring the surface-area normalized sorption capacity and the second-order rate constants of BPA uptake by the three sorbents.
  - The article provides new direct evidence for the partitioning of BPA into PUF through aqueous sorption, such as volumetric swelling of PUF fibrils induced by BPA uptake, thorough diffusion of BPA inside PUF fibrils, and hydrogen bonds between BPA and PUF functional groups.
    - The authors obtained this finding by using high-resolution electron microscopy, gas pycnometry, and cross-sectional infrared imaging analyses.
  - The article elucidates the specificity of PUF sorption by examining its sorption behaviors under different water chemistry, such as the effects of matrix compounds, sorption selectivity, alkaline pH sensitivity, and BPA recovery.
    - The authors obtained this finding by spiking BPA solutions with various water matrix constituents and measuring the changes in BPA uptake and desorption by PUF.
  - The article demonstrates the potential use of PUF as a superior sorbent medium for BPA in aqueous media, and suggests possible applications such as sorptive removal, extraction, and recovery of BPA from waste streams.
    - The authors obtained this finding by discussing the advantages of PUF over conventional adsorbents, such as high sorption capacity, rapid uptake rate, enhanced selectivity, easy regeneration, and good reusability.
- 

The article is an original research paper published in the *Journal of Materials Chemistry A*, which contains approximately 6,100 textual words, four figures, and one data table (Han et al. 2015). The study investigated the underlying mechanisms of the unusually high polymer-water partitioning coefficients of bisphenol A (BPA), a widely occurring plasticizer, in the presence of ether-type polyurethane foam, a consumer-grade elastomeric material, in aqueous media. The reported work addressed a specialized topic in the research domains of environmental analytical chemistry, water treatment, and polymer chemistry and physics.

To this end, we have tested the model further on three research papers of various types, lengths, and topics (Tables 3, 4 and 5). The first paper is a “Short Communication” article without a regular “Abstract”, “Conclusion”, or “Highlights”, making it impossible for the model to copy or paraphrase author-supplied text summaries from the paper (Han et al. 2013). As validated by the authors, the points listed in the model’s response contain details on the findings that are *not* mentioned in the synopsis, i.e., the one-sentence abstract or the figure/table captions, the only “shortcuts” for the model to access such information in the paper. This study was



among the first report of a series of experimental investigations on the penetrative diffusion and high-capacity accumulation, i.e., partitioning of trace organic contaminants, also referred to as “micropollutants” or “contaminants of emerging concerns”, into common plastics and elastomers in aqueous media, and their interactions on a molecule level. The second paper, published in *Talanta* in 2017, was a follow-up study of the two previous papers analyzed by the ChatGPT-enabled new Bing (Han et al. 2013, 2015). This paper was written in a lengthy and dense manner, with approximately 6600 textual words, six figures, and four data tables (Han et al. 2017a). The third paper is an original research article published in *Environmental Science & Technology*, which contains about 6000 textual words with four figures and one data table (Han et al. 2017b). The paper contains a short 200-word “Abstract” with no “Conclusion” or “Highlights”, as per the journal’s requirements. In this paper, we reported the accumulation and uncontrolled release of a broad-spectrum antibacterial (triclosan) in commercial toothpaste formulations in and from toothbrush bristles and head components, which attracted substantial interest from the press with more than 50 news reports in English-speaking countries. These news articles, all of which are available in the public domain, can help the model understand the research and thus provide even higher-quality answers to user’s queries.

In all three cases, the model consistently generated accurate and context-relevant answers by abstracting the key findings with information on the specific methods used by the authors for obtaining such findings. It should be reiterated that the responses generated by the model are not mere repetitions or rephrased words from the abstracts and conclusions, but deeper insights that require the model to delve into the full text to obtain. Using the bullet-point structure, the ChatGPT-enabled new Bing consistently generated succinct summaries on users’ queries, at language proficiency levels that sometimes surpass the original writing in terms of clarity and succinctness, a task that even well-versed scientists may find difficult. This implies that although the authors are most knowledgeable about their own research, ChatGPT can summarize the contents of their research papers *more effectively*. Such capability of large language models can be used for generating quick snapshots of one’s research work, e.g., for use in PowerPoint slideshows, three-minute thesis challenges, or academic social network sites. Indeed, ChatGPT can often outperform authors’ writing in terms of clear expression and conciseness in wording. The AI large language model managed to do this by understanding the *purpose* of the study and the *nature* of the results and leveraging its expert-level language skills to craft succinct statements and list them in a coherent manner. Most importantly, the model did it unthinkably fast. It would normally take 1–2 hours for a first-year Ph.D. student who already has some background knowledge on the research topic to get the nitty-gritty out of a dense research paper and come up with summaries that are as thorough and accurate as those in the examples. With ChatGPT, it is a matter of seconds. Even if the user needs to refine the prompt and ask further on certain points, the process would take no more than several minutes.

In conclusion, ChatGPT could boost the number of research papers that one could skim-read in a limited timeframe, and one could easily go deeper by asking questions on specific points that are *not* in the Abstract and the Conclusion of the paper, gaining

**Table 3** The key findings of a short communication article (Han et al. 2013) with information on how authors obtained these findings, using ChatGPT-enabled new Bing under the “Creative Mode”

**Prompt:** List the key findings in this article in detail with bullet points, with information on how authors obtained each of these findings.

**Response:** The key findings with bullet points in this article are:

- Polyamide 12 (PA12) microspheres showed remarkable sorption properties for triclosan (TCS), a broad-spectrum antibacterial agent and an endocrine disrupting contaminant in water.
  - The authors measured the sorption capacity, kinetics, selectivity and regenerability of PA12 for TCS in comparison with conventional porous sorbents such as activated carbon (AC) and polystyrene resin (PS-DVB).
- PA12 exhibited exceptional binding affinity for TCS in water, with a surface area-normalized sorption capacity of  $14.5 \text{ mg m}^{-2}$ , which is two orders of magnitude higher than AC and PS-DVB and the highest yet reported for any synthetic or natural sorbent.
  - The authors used nitrogen adsorption isotherms to determine the surface area and pore volume of the sorbents and used batch experiments to obtain the sorption equilibrium and kinetics data.
- PA12 showed extremely fast sorption kinetics for TCS in water, removing 98% of TCS from a  $1.0 \text{ mg L}^{-1}$  solution within 5 min, while AC and PS-DVB only removed 31% and 26%, respectively.
  - The authors attributed the fast sorption rate to the nonporous structure and small external surface area of PA12, which imposed no hindrance on TCS solute transport.
- PA12 displayed high sorption selectivity for TCS in water, showing much lower sorption capacities for structurally similar compounds such as TCS-methyl and phenol.
  - The authors also found that PA12 maintained consistent sorption capacities for TCS in different water matrices containing salt, phenol, or natural organic matter, while AC and PS-DVB showed notable deterioration due to competitive sorption and pore-blocking effects.
- PA12 was easily regenerated by contacting with alkaline solution, achieving complete desorption of TCS and high recovery of sorption capacity after multiple regeneration cycles.
  - The authors compared three different regeneration methods for AC and PS-DVB and found that they only yielded low to moderate regeneration efficiencies.
- The authors proposed a hybrid sorption mechanism for PA12 consisting of two steps: (1) adsorption of TCS solutes on PA12 surfaces via hydrophobic partitioning; (2) absorption of TCS molecules by the PA12 polymer matrix via strong hydrogen bonds on accessible amide groups.
  - The authors used quantum mechanical modelling to simulate the interactions between TCS molecules and PA12 amide groups in an aqueous environment and calculated the hydrogen bond distances and energies.

The two-level structure is added to the bullet points in the model’s response for clarity.

further insights into the research. Whether they are didactic short articles addressing topics of wide interests or lengthy research papers dealing with topics in specialized domains, the AI large language model consistently generated accurate, succinct, and context-relevant answers by extracting the key points and user-defined specific information from the full text of research publications.

The main findings and insights from this section are listed below:

- In this section, we demonstrate the use of AI, specifically a ChatGPT-enabled version of the Microsoft Bing search engine, to quickly extract key points and other user-requested specific information from research papers.

**Table 4** The key findings of a lengthy research paper (Han et al. 2017a) with information on how authors obtained these findings

---

**Prompt:** List the key findings in this article in detail with bullet points, with information on how authors obtained each of these findings.

---

**Response:** Here are the key findings with bullet points in this article in detail, with information on how authors obtained each of these findings.

- The authors investigated the specificity of polyurethane foam (PUF) as a solid-phase extractant for aqueous organic compounds, focusing on lipophilic phenolic compounds (LPCs) as a group of primary targets.
    - They used six bisphenol analogs and nine monofunctional LPCs as probing and validating compounds.
  - They found that molecular lipophilicity, hydrogen-bond acidity, and number of proton donors were the key molecular attributes that governed the extractability of LPCs by PUF.
    - They measured the partitioning coefficients ( $K_{PW}$ ) of the LPCs between PUF and aqueous phase and correlated them with their molecular properties using quantitative structure–property relationship calculations and density functional theory calculations.
  - They also found that PUF exhibited remarkable selectivity towards compounds with correlating molecular attributes against other structural analogs and co-existing matrix organics.
    - They demonstrated this by performing selective extraction experiments in mixed solutions and simulated water matrices containing various bisphenols (BPs) and other organic contaminants.
  - They proposed a qualitative assessment method to screen untested LPCs for PUF extraction using reference compounds and molecular criteria. They also suggested a conceptual model of PUF macromolecular structure as a flexible ‘molecular zipper network’ that is most affinitive towards nonionic, permeable and lipophilic guest molecules with multiple reactive proton donors.
- 

“Creative Mode” was used for generating the response using ChatGPT-enabled new Bing. Second-level bullet points and an explanation of the abbreviation “BPs” are added in the response for clarity.

- In all the cases we tested, the model delved into the full text and generated succinct and mostly correct responses. In addition to the key findings, users can obtain detailed information on how each of the findings was obtained in the study. This can be particularly useful for researchers who are interested in the methods used in the studies. Overall, these examples demonstrate the model’s ability to analyze specialized scientific content with varying article types, content structures, lengths, and research topics.
- The model’s ability to quickly extract accurate and pertinent information from scientific publications provides a valuable tool for researchers, particularly in broad or rapidly evolving research domains where there is a vast amount of literature and that staying updated with new evidence and findings is a challenge.

**Table 5** The key findings of a regular-length research paper (Han et al. 2017b) with information on how authors obtained these findings

---

**Prompt:** List the key findings in this article in detail with bullet points, with information on how authors obtained each of these findings.

---

**Response:** Some of the key findings with bullet points in this article are:

- **Triclosan (TCS)**, a widely used antimicrobial agent in toothpastes, can be **accumulated** by common toothbrush head components, such as **nylon bristles** and **thermoplastic elastomers (TPEs)**, during regular toothbrushing.
    - The authors obtained this finding by performing simulated brushing experiments with 22 types of commercial toothbrushes and six TCS-formulated toothpastes.
  - The amount and rate of TCS accumulation depend on the **material**, **design**, and **stiffness** of the toothbrush head components.
    - The authors obtained this finding by measuring the TCS uptake by different toothbrushes with various characteristics and comparing their accumulation profiles.
  - The accumulated TCS can be **gradually released** into toothpaste slurries after switching to TCS-free alternatives. The release rate varies with the type of toothpaste used, with faster release in peroxide-containing, alkaline, and surfactant-rich toothpastes.
    - The authors obtained this finding by eluting the TCS-laden toothbrushes with different TCS-free toothpastes and analyzing the TCS concentration in the eluates.
  - The accumulation effect is not exclusive to TCS but also applies to **other chemicals** present in TCS-TPs and a range of regular toothpastes, such as **flavoring agents**, **antioxidants**, and **UV filters**.
    - The authors obtained this finding by identifying the coextracted compounds in toothbrush eluates using liquid chromatography and gas chromatography-mass spectrometry.
- 

“Creative Mode” was used for generating the response using ChatGPT-enabled new Bing. In this example, the model used bold fonts to highlight the key information in the response, in an accurate manner.

### 3 Interpreting Figures and Correlating to Specific Conclusions

Figures often contain some of the most valuable information in scientific publications, but they can be difficult to understand as “stand-alone” features, i.e., without reading the main text. The amount of information carried by figures makes them “hot spots” for extracting essential information from research papers. However, some authors use symbols or abbreviations excessively with no interpretation in the figure caption explaining the data or trends in the figure, making them even more difficult to understand by speed readers.

Using ChatGPT or the model-enabled new Bing, one can *directly* analyze a figure in a research paper in the right context, regardless of whether the figure has been well prepared, e.g., in a “stand-alone” manner—a practice that we encourage our authors to do in their manuscripts. For example, when we tested the ChatGPT-enabled new Bing on two different types of figures, i.e., hyperspectral images with spectra (Han et al. 2015) and a set of data plots (Huang et al. 2020), the model provided accurate interpretation with enough details that would require one to read the figure along with the main text, rather than rephrasing the text from the figure captions.

### 3.1 *Hyperspectral Images with Spectra*

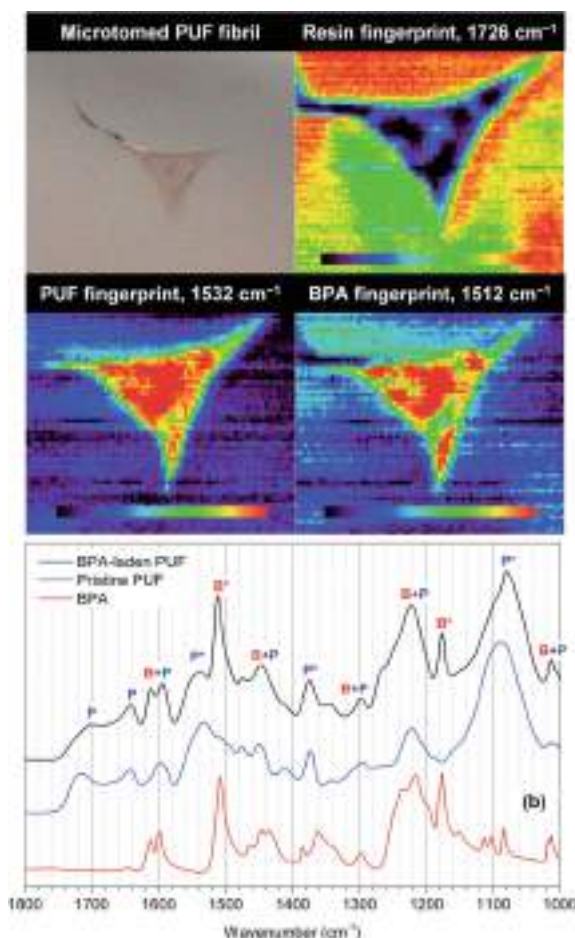
In the first example, we used a short prompt, i.e., “*Tell me in detail what Figure [x] shows and proves in this paper*” to ask new Bing what information the authors wanted to convey to readers in this particular figure. The original figure and caption are reprinted as references (Fig. 2). It is worthwhile to note that in the paper by Han et al. (2015), although the figure caption provides plenty of information on the sample preparation and data acquisition methods, it does *not* contain any interpretation of the results so that readers can quickly learn what conclusions are drawn from the evidence shown in this figure or the significance of the results at the time when this was published. After reading the model's response (Table 6), it is evident that the model delved into the full text and attempted to (1) summarize the authors' descriptions of the results shown in this figure; and (2) identify the specific conclusions in the paper that are supported by this particular figure.

The response generated by new Bing was validated as “mostly correct” by the original authors of this paper. Note that in the model's response, the color coding interpreted by the model was incorrect, which is marked by bold fonts in the response (Table 6). Also, the two spectral peaks should be “ $1697\text{ cm}^{-1}$ ” and “ $1080\text{ cm}^{-1}$ ”, or more accurately, “shifts from  $1715\text{ cm}^{-1}$  to  $1697\text{ cm}^{-1}$  and from  $1090\text{ cm}^{-1}$  to  $1080\text{ cm}^{-1}$ ” to align with the authors' discussions on the spectral data in the “Results and discussion” of the paper, under the subsection “Cross-sectional IR imaging analysis” (Han et al. 2015). The two approximated wavenumbers given by the model, i.e.,  $1720\text{ cm}^{-1}$  and  $1100\text{ cm}^{-1}$  which are not found in any part of the paper, may have been extracted by the model from the spectral data plots in the figure. This implies that ChatGPT is capable of “reading” graphical plots *directly* rather than relying on searches for relevant contents in the full text to interpret the figure.

### 3.2 *Data Plots*

We repeated this test by asking new Bing to analyze a set of data plots in a research paper by Huang et al. (2020). The original figure and figure caption are reprinted as reference (Fig. 3). The “More Balanced” mode, which generates “*Responses (that) are reasonable and coherent, balancing accuracy and creativity in conversation*”, was used for generating the response (Table 7). In this example, we used a more intuitive prompt, i.e., “*Please help me analyze the figures and tables in this article, and explain the information in each one in detail.*” With this prompt, the user essentially asked the model to explain *all* illustrations in this research paper. This is a challenging task because the paper contains three figures with Supplementary Data. Explaining each one of them would take time (even for the model) and yield a long response.

The model responded to the user's request by structuring its response in two distinct parts. The first part is a brief, step-by-step guide on analyzing figures and tables in research papers. This provides the general strategy for completing the task requested by the user. The second part is an example of the model's analysis of one



**Fig. 2** **a** Microscopy and infrared images of a microtomed cross-section of a bisphenol A (BPA)-laden polyurethane foam (PUF) fibril embedded in a resin medium. Infrared images were constructed from hyperspectral data using the fingerprint peaks of LR White resin, polyurethane foam (PUF), and bisphenol A (BPA) selected from their major infrared absorption bands with minimal interferences. Color scale indicates the relative intensity of infrared absorbance which corresponds to the abundance of the material or compound. Field of view in each infrared image displays  $75 \times 75$  pixels at a resolution of  $1.56 \times 1.56 \mu\text{m}$  per pixel. **b** Infrared spectra of bisphenol A (BPA), pristine polyurethane foam (PUF), and bisphenol A (BPA)-laden polyurethane foam (PUF). The bisphenol A (BPA) analyzed was in solid form with a purity of  $\geq 99\%$ . The spectrum of bisphenol A (BPA)-laden polyurethane foam (PUF) was extracted from the infrared image of a pretreated polyurethane foam (PUF) fibril with bisphenol A that partitioned into the fibril accounting for approximately 40% of the dry fibril weight. The spectra of pristine and bisphenol A (BPA)-laden PUF were normalized based on aromatic out-of-plane C–H bending peak at  $928 \text{ cm}^{-1}$  and  $\text{CH}_2$  stretching peak at  $2871 \text{ cm}^{-1}$  which were inert to interactions with bisphenol A (BPA) and had minimal interference from its infrared absorption bands. Infrared absorption bands of BPA (B) and PUF (P) are labeled in the spectrum where the asterisk denotes the dominant contributor. Reproduced from Han et al. (2015) with permission from the Royal Society of Chemistry

**Table 6** Interpreting a figure in a research paper and correlating the figure to specific conclusions in the paper

---

**Prompt:** Tell me in detail what Figure 3 shows and proves in this paper.

---

**Response:** Based on the web page context and the web search results, Fig. 3 shows and proves the following points:

- Figure 3 is a cross-sectional infrared image of a PUF fibril that has absorbed BPA from aqueous media. The image shows the distribution of **BPA (red)** and **PUF (green)** along the fibril cross-section.
  - Figure 3 proves that BPA can diffuse and partition into PUF through aqueous sorption, as evidenced by the presence of BPA inside the PUF fibril and not just on the surface.
  - Figure 3 also proves that BPA can form hydrogen bonds with PUF carbonyl and ether oxygen groups, as indicated by the spectral peaks at **1720 cm<sup>-1</sup>** and **1100 cm<sup>-1</sup>**, respectively. These hydrogen bonds may enhance the affinity and selectivity of PUF for BPA.
  - Figure 3 provides direct evidence for the solid-phase partitioning mechanism of PUF sorption for BPA, which differs from the conventional surface adsorption mechanism of other porous adsorbents.
- 

The “Figure 3” mentioned in the user prompt and the model’s response refers to the original numbering of this figure in the referenced study (Han et al. 2015). Readers may refer to the full text of the paper for more information and further validation.

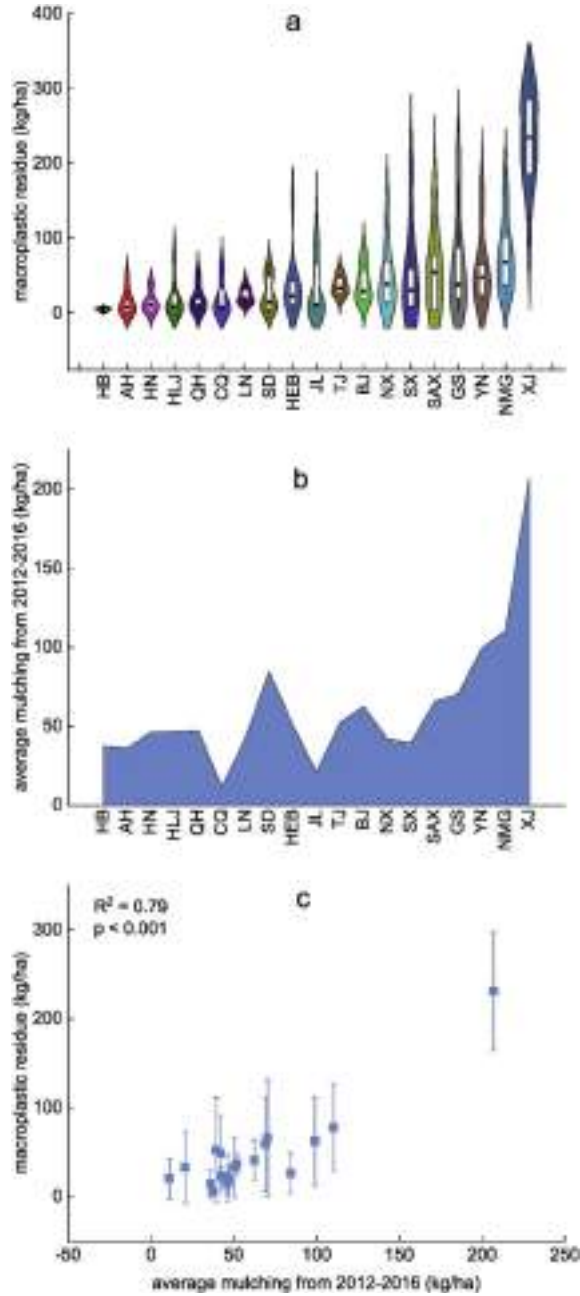
particular figure in this paper (reprinted in Fig. 3). The second part of the response provides three answers: (i) a succinct summary of the results shown in this figure, i.e., *what the figure shows*, (ii) what authors said about this figure in their paper, i.e., *author’s interpretation*; (iii) what specific argument this figure supports in this paper, i.e., *what the figure proves*.

There are several commendable points in the model’s response (Table 7). First, the model structured its response in two parts. This is a smart strategy because, in response to the user’s demanding request, the model essentially told the user: “This is a big ask. But you can do this too. I’ll tell you how you could do it, step by step. And here is an example that I did.” This is very intelligent and human-like when someone, e.g., presumably an expert on the question being asked, is faced with a “big question” from the audience with limited time to answer and he or she is still trying to be very helpful. Second, if we take a closer look at the answer, the first part of the response is a high-quality mini-tutorial that can be readily taught in class for college students. The comments by the model on the third and the last bullet points in the first part of the response, i.e., “*The text should provide context and highlight the main findings or implications of the data. The text should also avoid repeating information that is already shown in the figure or table.*” and “*Evaluate how well the figure or table supports or illustrates the main argument or purpose of the article*” revealed the deeper links between figures, discussion texts, and conceptualization of the research. They taught us a good lesson on how figures should be used in research papers to communicate the “full picture” of scientific discovery to readers more effectively.

In the example shown in the second part of the response, the model provided a succinct summary of the results shown in the figure, located the authors’ discussion of the figure in the paper, excerpted the relevant texts, and attributed a main argument in this paper to the evidence in this figure, all in an *accurate* and *concise* manner. Notably, the first point of the answer contains data and information that



**Fig. 3** Relationship between macroplastic residues and the use of plastic mulching in agricultural soils across China: **a** violin plots of the abundances of macroplastic residues in agricultural soils across China, **b** average mass of mulching film in 2012–2016, **c** relationship between macroplastic residues and the use of plastic mulching film. Reproduced from Huang et al. (2020) with permission from Elsevier B.V.





**Table 7** Analyzing data plots in a research paper by Huang et al. (2020)

**Prompt:** Please help me analyze the figures and tables in this article, and explain the information in each one in detail

**Response:** The response is from both the web page context and the web search results.

To analyze the figures and tables in this article, you need to understand their purpose, content, format and relation to the text. Here are some steps you can follow:

- Identify the type of figure or table (e.g., graph, chart, map, diagram, photo) and its title. The title should be descriptive and informative about what the figure or table shows.
- Examine the data or information presented in the figure or table. Look for patterns, trends, outliers, comparisons, relationships, or differences among the data. Pay attention to any labels, units, scales, legends, keys, or captions that explain how to interpret the data.
- Read the text that refers to or explains the figure or table. The text should provide context and highlight the main findings or implications of the data, The text should also avoid repeating information that is already shown in the figure or table.
- Evaluate how well the figure or table supports or illustrates the main argument or purpose of the article. Consider if it is clear, accurate, relevant consistent and appropriate for its intended audience.

For example:

Figure 1 shows a violin plot of microplastic residues in agricultural soils across China. It compares 19 provinces based on their average abundance of macroplastics (kg/ha) in mulching farmlands. The plot shows that Xinjiang has by far the highest abundance of macroplastics (mean: 230.9 kg/ha), followed by Inner Mongolia (mean: 77.6 kg/ha) and Gansu Province (mean: 65.9 kg/ha). The plot also shows that most provinces have a skewed distribution of macroplastics, with some outliers having much higher values than others.

The text that refers to Fig. 1 is on page 2 of this article. It states that “a national survey of macroplastic residues (> 5 mm) in mulching soils across China was conducted” and that “a highly significant linear correlation is observed” between macroplastic residues and plastic mulching film use.

Figure 1 supports one of the main arguments of this article: that plastic film mulching is a major source of macroplastics in farmlands in China. It provides visual evidence for this claim by showing how different provinces vary in their macroplastic abundance depending on their plastic mulching film consumption.

The “More Balanced” mode, which generates “Responses (that) are reasonable and coherent, balancing accuracy and creativity in conversation”, was used in this example. In the response, “Figure 1” refers to the original numbering in the referenced study. Note the instructions given by new Bing on analyzing figures and tables in research papers, particularly the third and fourth points, which are useful for both interpreting and preparing figures in research papers. In this example, the user entered a prompt that would require extensive answers from the model. In response to this request, the model provided a general strategy upfront, followed by the analysis of one particular example, a common strategy that we have found on other tests on the AI large language model. For users, it is more appropriate to specify one or two figures for the model to analyze, e.g., “Analyze figure [x] in this paper and explain it in detail” or “Tell me what Figure [x] shows and proves in this paper”, as we have done in the previous example.